面向21世纪课程教材
Textbook Series for 21st Century

分子遗传学

杨业华 主编

中国农业出版社

面向21世纪课程教材
Textbook Series for 21st Century

分子遗传学

杨业华　主编

中国农业出版社

图书在版编目（CIP）数据

分子遗传学/杨业华主编．－北京：中国农业出版社，2000.11

面向21世纪课程教材

ISBN 7－109－06635－5

Ⅰ．分… Ⅱ．杨… Ⅲ．分子遗传学－高等学校－教材 Ⅳ．Q75

中国版本图书馆CIP数据核字（2000）第53550号

中国农业出版社出版

（北京市朝阳区农展馆北路2号）

（邮政编码100026）

出版人：沈镇昭

责任编辑 江社平

中国农业出版社印刷厂印刷 新华书店北京发行所发行

2001年1月第1版 2001年1月北京第1次印刷

开本：787mm×960mm 1/16 印张：34.75 插页1

字数：627千字 印数：1～10 000册

定价：45.70元

（凡本版图书出现印刷、装订错误，请向出版社发行部调换）

前　言

遗传学是生物科学的基础学科，它发展迅速，包容知识面广，理论性强且又与实际应用联系密切。遗传学课程是高等农林院校动物、植物和微生物类各专业重要的专业基础课程，为了使遗传学教学更好地适应本学科的迅速发展和密切配合相关的专业课程教学，亟须编写出一套能系统阐述遗传学基本原理、反映现代遗传学水平、追踪学科发展前沿的遗传学教材。

根据教育部“高等教育面向21世纪教学内容和课程体系改革”精神并结合高等农林院校生物类各学科特点，我们编写了《普通遗传学》和《分子遗传学》两本教材，以期将遗传学分成两个教学板块，分层次、按专业特点和要求教学。这里所说的“普通遗传学”实为基础遗传学，它讲述除分子遗传学内容以外的遗传学各分支学科的一般原理及常规分析方法，适合为生物类各专业本科学生奠定遗传学基础。第二板块“分子遗传学”讲授第一板块中除数量性状遗传和群体遗传以外的各分支学科的深层次内容，重点在基因的结构、功能、调节表达、基因突变、遗传信息复制以及基因操作等内容，适合对遗传学知识要求较全面的专业及部分对遗传学有特殊兴趣的学生进一步学习。《普通遗传学》由高等教育出版社出版。

在编写思想上，我们力求本系列教材便于学生自学、有利于调动学生学习自主性、有利于培养学生自我获取知识的能力、有利于启发学生的创造性思维。

在编写内容上，《普通遗传学》除了系统阐述遗传学的经典理论和基本原理外，还全面充实了植物、动物和微生物方面遗传分析方法学内容，以拓展学生知识面和加强学生分析问题和解决问题基本能力的培养。《分子遗传学》重点论述基因的特征、调节表达机理以及基因操作的基本原理，引进了遗传学和分子生物学研究的新成果和新概念，使学生在修完本课程后，能够把握各类生物遗传信息表达的共性和特点，把握学科发展的总体趋势，为进一步自学奠定较全面的理论基础。

《分子遗传学》分九章。第一章介绍DNA和RNA作为遗传物质的特征和

遗传物质损伤的修复机制，以及遗传信息表达的基本过程。第二章介绍基因的特征，包括基因的结构特征、基因的组织、基因突变与重组的分子基础、转座因子及其转座的遗传效应。第三章介绍原核生物基因组的复制、基因转录与翻译调节。第四章介绍真核生物基因组的组织及基因的调节表达，包括染色体的特征、DNA复制、基因转录和翻译启始调节、转录子和蛋白质的加工，发育阶段及环境信号对基因表达的调节等。第五章介绍真核生物细胞器线粒体和叶绿体基因组的特征，基因的调节表达，细胞器基因组与核基因组的互作等内容。第六章主要介绍各种生物代表性病毒基因组的结构特征、DNA复制和基因表达的策略以及病毒侵染对宿主细胞的影响。第七章介绍脊椎动物免疫的基因控制，重点介绍抗体基因重组与抗体多样性的分子基础。第八章介绍病毒、细菌和真核生物中模式生物的基因对发育和细胞分裂和生长的控制。第九章介绍基因操作的基本原理及其应用，包括各类载体的基本特征及基因重组的方法，动植物遗传转化方法及其原理，细菌、病毒表达系统在工农业、商业、医学方面的应用等内容。

《分子遗传学》按80学时编写，生物类各专业可根据本专业特点和教学时数，对讲授内容作适当取舍。本书每章之后都有内容小结和习题，便于学生复习、巩固和把握重点。

本教材是教育部“高等农林院校面向21世纪生物系列课程教学内容和课程体系改革”项目成果之一，亦得到中华农业科教基金部分资助。华中农业大学杨业华编写第一至第五章，毕丁仁编写第六、七章，赵书红编写第八章，中国农业大学王国英和湖北大学金山编写第九章，南京农业大学张红生编写了第八章部分内容。限于我们的水平，书中舛误之处在所难免，特别是在本教材的内容选取方面，是否切合21世纪农林专门人才培养的需要，尚有待时间检验，诚望遗传学家和广大读者批评指正。

本书初稿完成后，由中国农业大学陈永福教授主审全书，西南农业大学裴炎教授亦审阅了全书，在此致以诚挚感谢。

编　者

1999年12月

主　　编　杨业华（华中农业大学）
副 主 编　毕丁仁（华中农业大学）
　　王国英（中国农业大学）
参编人员　金　山（湖北大学）
　　张红生（南京农业大学）
　　赵书红（华中农业大学）

主　　审　陈永福（中国农业大学）
参　　审　裴　炎（西南农业大学）

内容简介

本书系统阐述了遗传信息的复制，基因的结构、功能和调节表达机理，基因突变与重组，基因对发育的控制以及基因操作及其应用的基本原理。全书分九章，包括遗传物质DNA和RNA的特性、基因的特征、原核生物基因组及其表达、真核生物细胞核基因组及其表达、线粒体和植物叶绿体基因组及其表达、病毒基因组及其表达、脊椎动物免疫的分子基础、基因对发育和细胞周期的控制、基因操作及其应用。为便于读者学习，每章之后编有提要和复习题，书末编有中、英文索引和主要参考书。

本书是教育部“面向21世纪课程教材”之一，可作为高等农林院校生物类各专业本科生的遗传学基础课教材，亦可作为综合性大学和高等师范院校生物类有关专业本科生教材或教学参考书以及供有关专业的研究生和科技工作者参考。

目　录

前　言

第一章　遗传物质 DNA 和 RNA …… 1

第一节　DNA 分子的结构与复制 …… 1

一、遗传物质 DNA 的发现与证明 …… 1

二、DNA 分子的结构 …… 4

三、DNA 分子的功能单位 …… 9

四、DNA 分子的复制 …… 10

五、DNA 序列组成 …… 13

第二节　RNA 的结构与功能 …… 15

一、RNA 分子的结构 …… 15

二、RNA 合成与病毒 RNA 遗传信息传递 …… 16

三、细胞中主要的 RNA …… 18

第三节　三联体密码与蛋白质合成 …… 22

一、三联体密码 …… 22

二、翻译 …… 25

三、中心法则及其发展 …… 28

第四节　DNA 损伤及其修复 …… 29

一、DNA 变异 …… 30

二、DNA 损伤的修复机制 …… 32

小结 …… 39

复习题 …… 40

第二章　基因的特征 …… 42

第一节　基因的结构特征 …… 42

一、基因的一般结构 …… 42

二、基因的变异类型 …… 46

第二节　基因的组织 …… 51
一、操纵元 …… 51
二、重复基因与基因家族 …… 53
三、重叠基因 …… 56
第三节　基因突变与重组 …… 57
一、基因突变 …… 57
二、基因重组 …… 68
三、基因转变 …… 73
第四节　转座子 …… 76
一、转座子的遗传特征 …… 76
二、细菌中主要的转座子 …… 77
三、植物转座子 …… 81
四、酵母的转座子 …… 83
五、无脊椎动物的转座子 …… 85
六、转座子转座与基因突变 …… 88
小结 …… 94
复习题 …… 96

第三章　原核生物基因组及其表达 …… 100

第一节　细菌基因组及其复制 …… 100
一、细胞中的拟核 …… 100
二、细菌 DNA 复制的主要酶类 …… 102
三、细菌 DNA 复制的模型 …… 107
四、质粒的一般特征及其复制调节 …… 112
第二节　细菌基因组的转录与翻译 …… 114
一、与转录有关的酶 …… 114
二、转录启始、延长与终止 …… 116
三、翻译启始 …… 123
第三节　操纵元的调节表达 …… 129
一、操纵元调节表达的一般模型 …… 129
二、乳糖操纵元的负控制和正控制诱导表达 …… 131
三、色氨酸操纵元与弱化反应 …… 133
四、核糖体蛋白操纵元与翻译自动调节 …… 136
五、转录的总体控制 …… 139
六、热激蛋白的合成 …… 140
小结 …… 141

复习题 …… 144

第四章　真核生物细胞核基因组及其表达 …… 146

第一节　DNA 与染色体 …… 146
一、组蛋白与染色质 …… 146
二、着丝粒与卫星 DNA …… 149
三、端粒 …… 150
四、分散的高度重复 DNA …… 151
五、细胞分裂中期的骨架 …… 152
第二节　细胞核 DNA 复制 …… 152
一、复制子及复制启始区 …… 152
二、DNA 复制的蛋白质或酶类 …… 154
三、细胞核 DNA 复制模型 …… 156
四、染色体端粒 DNA 的合成 …… 158
五、有选择的 DNA 复制与基因扩增 …… 158
第三节　基因的转录与蛋白质合成 …… 162
一、转录酶 …… 162
二、转录的控制信号 …… 163
三、转录启始及转录因子 …… 169
四、mRNA 的加工 …… 181
五、蛋白质合成的启始 …… 187
六、翻译延长和终止 …… 190
第四节　细胞核基因表达的控制 …… 191
一、基因表达在转录和转录后水平上的控制 …… 191
二、基因表达在翻译水平上的调节 …… 197
三、基因表达的翻译后调节 …… 201
第五节　发育阶段对基因表达的调节 …… 207
一、基因表达的细胞、组织和器官特异性 …… 207
二、脊椎动物珠蛋白基因表达的发育阶段调节 …… 209
三、性别对基因表达的调节 …… 210
第六节　环境信号对基因表达的调节 …… 214
一、半乳糖对酵母基因表达的调节 …… 214
二、外源激素对基因表达的调节 …… 216
三、热激效应对基因表达的调节 …… 217
四、光照对植物基因表达的调节 …… 219
五、缺氧对植物基因表达的调节 …… 220

小结 …… 221
复习题 …… 226

第五章 线粒体和叶绿体基因组的组织及表达 …… 228

第一节 线粒体基因组 …… 228
一、线粒体 DNA 的复杂性 …… 228
二、哺乳动物线粒体基因组 …… 230
三、酵母线粒体基因组 …… 231
四、植物线粒体基因组 …… 233
五、原生动物线粒体基因组 …… 236
六、线粒体基因密码的不通用性 …… 237
七、线粒体基因组的复制 …… 238
第二节 线粒体基因的转录与翻译 …… 238
一、线粒体基因的转录和转录控制 …… 238
二、线粒体转录子的加工 …… 239
三、RNA 编辑 …… 242
四、线粒体 mRNA 的翻译和翻译控制 …… 245
第三节 植物质体基因组的组织及表达 …… 247
一、质体基因组的组织和结构 …… 247
二、质体基因的表达 …… 251
第四节 线粒体和质体基因组与细胞核基因组的互作 …… 255
一、细胞器的生物发生与核基因组的关系 …… 255
二、细胞溶质蛋白向细胞器的运输 …… 256
三、细胞器基因与核基因表达的协调性 …… 260
四、植物线粒体与细胞质雄性不育性 …… 261
小结 …… 262
复习题 …… 263

第六章 病毒基因组的表达及病毒感染对宿主细胞的遗传效应 …… 265

第一节 病毒基因组的组织 …… 265
一、双链 DNA 噬菌体基因组 …… 265
二、哺乳动物双链 DNA 病毒基因组 …… 270
三、植物双链 DNA 病毒基因组 …… 273
四、单链 DNA 病毒基因组 …… 273
五、RNA 病毒基因组 …… 275

六、逆转录病毒基因组 …… 281
七、植物类病毒和拟病毒基因组 …… 284
第二节　病毒基因组的复制 …… 284
一、线状双链 DNA 病毒基因组的复制 …… 284
二、环状双链 DNA 病毒基因组的复制 …… 287
三、单链 DNA 病毒基因组的复制 …… 290
四、单链 RNA 病毒基因组的复制 …… 293
五、逆转录病毒基因组的复制 …… 293
第三节　病毒基因的转录及蛋白质合成 …… 298
一、病毒基因的转录启始信号 …… 298
二、病毒基因的转录终止信号 …… 306
三、RNA 病毒蛋白质合成的策略 …… 308
第四节　病毒与肿瘤 …… 311
一、肿瘤及癌的一般生物学概念 …… 311
二、动物及人体的致癌因子 …… 312
三、癌基因、抑癌基因及肿瘤转移调节基因 …… 316
四、DNA 肿瘤病毒与癌 …… 323
五、RNA 病毒与肿瘤 …… 327
第五节　病毒感染与宿主干扰素的合成 …… 329
一、干扰素 …… 329
二、干扰素基因 …… 331
三、干扰素合成 …… 333
四、干扰素的抗病毒活性 …… 334
小结 …… 336
复习题 …… 338
第七章　脊椎动物免疫的基因控制 …… 339
第一节　脊椎动物的免疫反应 …… 339
一、免疫的概念 …… 339
二、免疫系统与免疫反应 …… 341
三、抗原及抗原性 …… 343
第二节　B 细胞免疫反应 …… 345
一、抗体、单克隆抗体及其基因工程抗体 …… 345
二、抗体基因及其重组 …… 350
三、抗体基因重排与抗体多样性 …… 355
第三节　T 细胞免疫反应 …… 361
一、T 细胞的类型及对外源抗原的识别 …… 361
二、T 细胞受体基因 …… 363

三、TCR 基因的重排及其多样性产生的机制 …… 366
第四节 主要组织相容性复合体 …… 367
一、小鼠主要组织相容性复合体 …… 367
二、人类 MHC－HLA 复合体 …… 369
三、MHC 的生物学意义 …… 371
小结 …… 373
复习题 …… 376

第八章 基因对发育和细胞分化的控制 …… 378

第一节 基因控制发育的方式 …… 378
一、发育阶段的时间决定 …… 378
二、基因控制发育的方式 …… 381
三、Wnt 基因家族与发育 …… 382
第二节 λ噬菌体的发育调节 …… 387
一、λ噬菌体生长概况 …… 387
二、CⅠ抑制子和 Cro 蛋白的作用机制 …… 391
三、CⅡ蛋白与噬菌体发育 …… 392
四、整合酶基因表达的逆调节 …… 393
五、N 和 Q 蛋白的反终止作用对 λ 噬菌体发育的正控制 …… 394
六、噬菌体 DNA 复制的特殊启动因子 …… 399
七、单位 λ 基因组的形成 …… 400
八、λ噬菌体裂解周期的诱导 …… 400
第三节 细菌发育的基因控制 …… 400
一、细菌芽孢形成的基因控制 …… 400
二、鼠沙门氏杆菌表面抗原的转换 …… 403
第四节 酵母交配型的转换 …… 404
一、酵母的生活周期 …… 404
二、酵母的快速性转变 …… 405
三、酵母交配型盖合转换的调节 …… 407
四、mat a 和 mat α 基因编码控制蛋白 …… 409
五、外激素与单倍体细胞的交配感受性 …… 411
第五节 锥虫表面抗原转换的基因控制 …… 411
一、锥虫生活史及表面抗原转换现象 …… 411
二、锥虫表面抗原转换的机理 …… 413
第六节 高等真核生物发育阶段的基因控制 …… 417
一、果蝇中编码体型决定因子的基因 …… 417
二、果蝇的 toll 基因与形态发生梯度 …… 422
三、果蝇器官分化的同源异型基因 …… 422

四、分化的反式决定 …… 426
五、细胞质定子 …… 429
六、植物花芽分化的同源异型基因 …… 431
第七节　基因对细胞周期的控制 …… 433
一、细胞周期事件与基因表达 …… 433
二、细胞生长的控制 …… 440
三、细胞程序死亡 …… 445
小结 …… 446
复习题 …… 449

第九章　基因操作技术及其应用 …… 451

第一节　DNA 重组的基本技术 …… 451
一、核酸酶和限制性核酸内切酶与 DNA 分子的体外切割 …… 452
二、限制性片段长度多态性 …… 455
三、限制片段作图 …… 456
四、限制性片段的连接与克隆 …… 459
五、DNA 序列分析 …… 468
六、聚合酶链式反应 …… 472
第二节　基因操作的载体系统 …… 474
一、质粒载体 …… 474
二、单链噬菌体克隆载体 …… 476
三、噬菌粒——含有丝状噬菌体复制起始位点的质粒 …… 477
四、双链噬菌体载体 …… 478
五、黏粒载体 …… 480
六、酵母人工染色体和细菌人工染色体 …… 481
七、表达载体 …… 483
第三节　酵母的遗传工程 …… 486
一、酵母载体 …… 486
二、酵母载体的应用 …… 488
第四节　植物的遗传转化 …… 492
一、转基因植物 …… 492
二、植物基因操作的研究方向与重要基因的鉴定 …… 497
第五节　动物的遗传转化 …… 504
一、动物转基因的方法 …… 505
二、转化体的鉴定和分析 …… 505
三、重组 DNA 分子导入受体细胞 …… 505

四、动物基因工程的发展前景 …… 506
五、动物体细胞克隆技术 …… 507
第六节　重组 DNA 技术在医学上的应用 …… 508
一、基因治疗 …… 508
二、爱滋病的治疗 …… 510
三、生物技术新药 …… 512
四、基因工程与疫苗生产 …… 513
五、重组 DNA 探针与遗传病诊断 …… 516
小结 …… 518
复习题 …… 521

中文索引 …… 522
英文索引 …… 531
主要参考书 …… 540

第一章

遗传物质 DNA 和 RNA

作为传递遗传信息的物质，它首先必须能够保证物种的连续性和适应进化的需要，因此遗传物质必须含有生物学上有用的信息，这些信息必须能够稳定复制和稳定传递。其次，遗传物质必须能够自我表达，产生其他一些生物学分子乃至细胞和生物体，也就是说，必须要有某种机制将遗传物质中所含信息翻译成产品。此外，遗传物质还必须能够发生变异，因为生物进化要求遗传物质能够变异。自从埃弗里（Avery，1944）证明脱氧核糖核酸（deoxyribonucleic acid，DNA）是一种能够将遗传信息从一个细菌细胞传递到另一个细菌细胞的物质以后，又发现某些病毒中核糖核酸（ribonucleic acid，RNA）也是遗传信息的载体。经过大量深入研究，现在人们对 DNA 和 RNA 分子的生物学特征已有了充分认识。DNA 和 RNA 分子的结构非常有规律，非常适合贮存遗传信息，细胞内有关的酶能够对贮存的信息进行复制、修复和阅读，所以 DNA 和 RNA 具备遗传信息载体的所有特征。

第一节　DNA 分子的结构与复制

一、遗传物质 DNA 的发现与证明

（一）细菌的转化

研究遗传物质的实验可以追溯到 1928 年格里菲斯（Griffith，F.）利用肺炎双球菌（*Diplococcus pneumoniae* 或 *Pneumococcus*）感染家鼠的试验。肺炎双球菌有两种类型，一种是光滑型（S 型），被一层荚膜多糖所保护，具有毒性，在培养基上形成光滑菌落。另一种为粗糙型（R 型），无荚膜和毒性，在培养基上形成粗糙型菌落。S 型和 R 型还可按血清免疫反应不同，分成许多抗原型，如 SⅠ、SⅡ、SⅢ、RⅠ、RⅡ等。

S 型细菌感染家鼠后可使家鼠患肺炎而死亡。这种致病性是由其细胞外层的荚膜多糖（capsular polysaccharide）决定的，因为这种多糖可使细菌免遭宿

主细胞的破坏。S 型细菌经热处理（65℃）被杀死后就丧失感染动物的能力，但是将高温杀死的 S 型细菌与无毒的活的 R 型细菌混在一起注射到家鼠体内后，可使家鼠患肺炎而死亡（图 1－1），而从家鼠尸体中回收到的细菌又是有

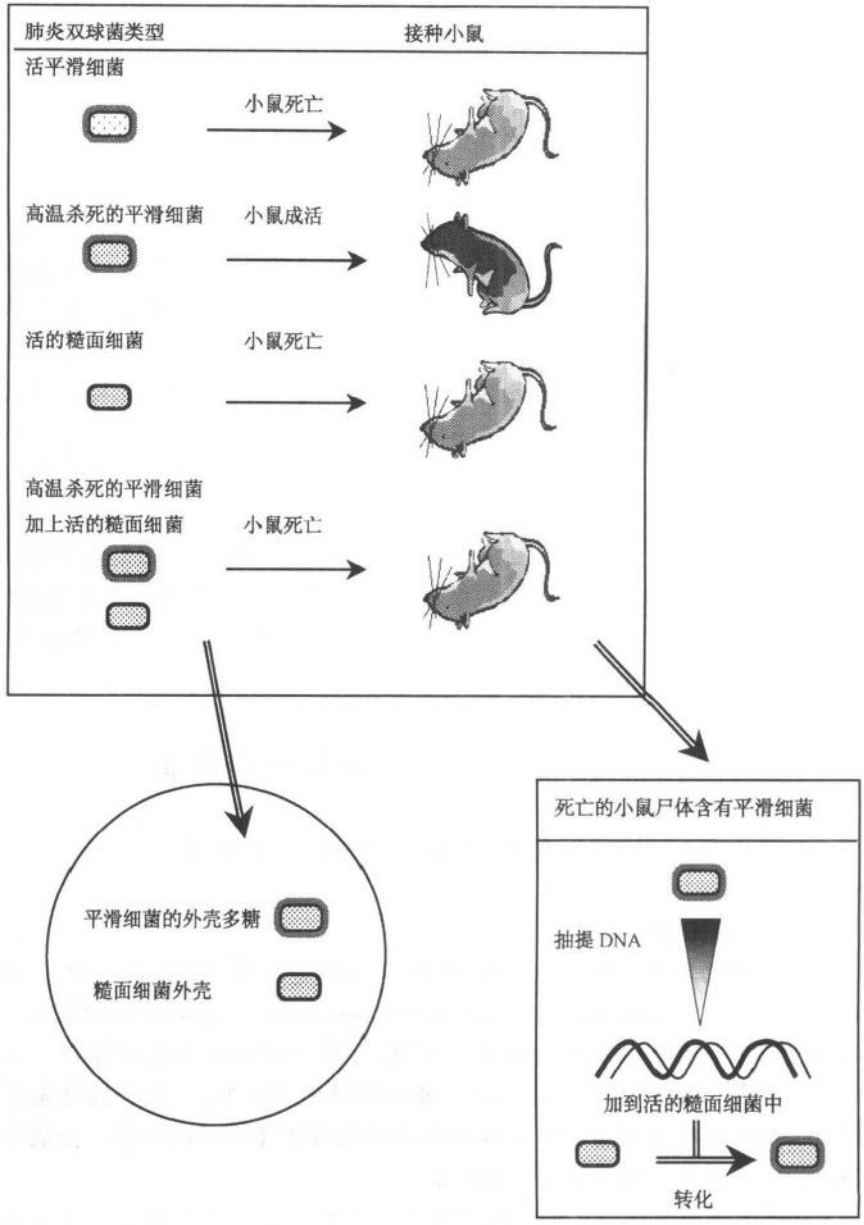


图 1－1　肺炎双球菌感染家鼠的实验

毒的S型细菌。

在这一实验中，灭活的S型细菌为SⅢ型，活的R型细菌为RⅡ型。从上述混合感染试验中回收到的有毒细菌为SⅢ型。因此杀死的SⅢ型细菌可使活的RⅡ型细菌合成SⅢ型荚膜多糖而成为有毒细菌，这种现象叫做转化(transformation)。

当时并不知道死亡细菌中与转化有关的物质是DNA，于是就把这种物质叫做转变因素(transforming principle)。尽管在这以前几十年就已知道真核生物的染色体由DNA和蛋白质组成，但直到1944年，埃弗里(Avery，T.)等才证明这种转变因素是脱氧核糖核酸(DNA)。他们不仅成功地重复了上述实验，而且将SⅢ型细菌的DNA加到RⅡ型细菌培养物中，也成功地使少数RⅡ型细菌转化为SⅢ型细菌。

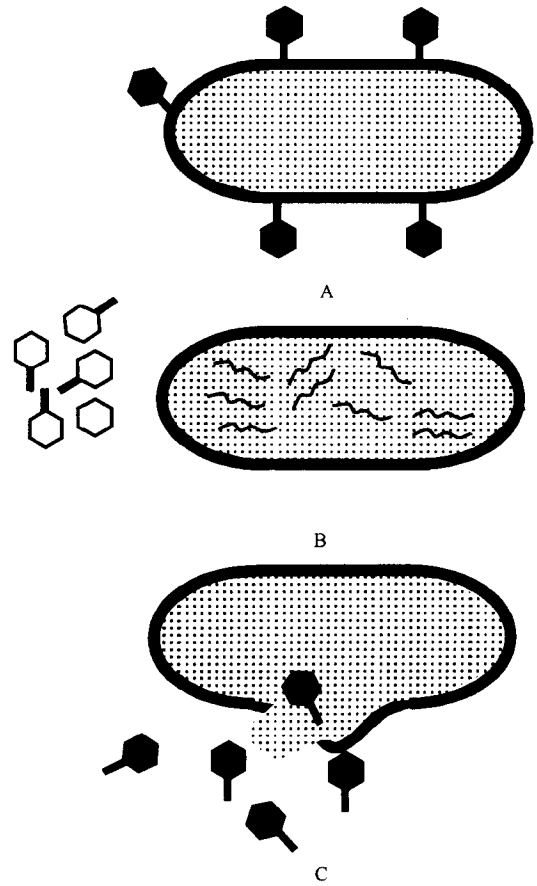


图1－2 赫尔希等证明DNA是遗传物质的实验

A. 用分别以^{32}P标记DNA和^{35}S标记蛋白质的噬菌体感染细菌 B. 分离噬菌体外壳和被感染的细菌，噬菌体外壳含有80%的^{35}S标记，被感染的细菌含有70%^{32}P标记 C. 分离子代噬菌体粒子，子代噬菌体有30%的^{32}P标记，只有1%以下的^{35}S标记

(二) 噬菌体的侵染与繁殖

噬菌体为一类细菌病毒，其中有一种T2噬菌体侵染大肠杆菌(*Escherichia coli*)。将噬菌体加到大肠杆菌中后，它们便吸附到细菌细胞的外表，将外壳内的一部分物质注入细胞内。大约在20min后，细菌细胞裂解，释放出大量噬菌体后代颗粒(图1－2)。

1952年，赫尔希(Hershey，A.D.)等用同位素^{32}P和^{35}S分别标记T2噬菌体的DNA和蛋白质，然后再分别用具有不同标记的噬菌体侵染大肠杆菌。约10min后，通过搅拌和离心，分离出两种成分。

一种成分为细菌本身，另一种成分为噬菌体的外壳。实验表明，基本上全部^{32}P的放射性见于细菌内，而且细菌被裂解后释放出来的噬菌体仍具有放射性，表明这种放射性可以传递给子代。由于噬菌体外壳由蛋白质组成，所以具有^{35}S的放射性，但这种放射性不能传给子代。这一实验再次证明了遗传物质是 DNA 的普遍性结论。

细菌和噬菌体的遗传物质是 DNA，那么真核生物的遗传物质又是否是 DNA 呢？在相当长一段时间内，人们只是一种推测。现在已有各种方法将外源 DNA 引入真核细胞内，产生新的遗传重组类型。例如，通过微量注射，将不同生物的 DNA 引入家鼠卵细胞中，这种 DNA 可以成为家鼠的一部分能够稳定遗传的物质（见第九章）。这种实验不仅直接证明了 DNA 是真核生物的遗传物质，而且还说明 DNA 可以在不同物种之间进行转移并保持功能。

虽然所有已知生物和许多病毒的遗传物质是 DNA，但是某些病毒如烟草花叶病毒等则使用另一种核酸即核糖核酸（RNA）作为遗传物质。RNA 分子在结构上只与 DNA 有微小差别，在功能上也非常相似。总的说来，所有遗传物质都是核酸，DNA 是主要遗传物质，在某些病毒中 RNA 也是遗传物质。

二、DNA 分子的结构

（一）DNA 分子的化学组成

DNA 分子是一种高分子化合物，由 4 种相似、但又不相同的核苷酸组成。每一个核苷酸包括一个磷酸基团，一个脱氧核糖和一个含氮杂环碱（图 1－3）。这 4 种核苷酸中的碱基分别是腺嘌呤（adenine，A）、鸟嘌呤（guanine，G）、胸腺嘧啶（thymine，T）和胞嘧啶（cytocine，C）。腺嘌呤和鸟嘌呤都是双环碱，称为嘌呤碱（purine），胸腺嘧啶和胞嘧啶为单环碱，称嘧啶碱（pyrimidine）。这些核苷酸的脱氧核糖的3'位和 5'位的碳原子通过磷酸二酯键连接起来，形成 DNA 分子。

脱氧胞嘧啶核苷酸（dCMP）　　脱氧鸟嘌呤核苷酸（dGMP）

脱氧胸腺嘧啶核苷酸（dTMP）　　脱氧腺嘌呤核苷酸（dAMP）

DNA 分子是一种极性分子。在 DNA 分子的一端即末端核苷酸的第 5 位碳原子上有一个磷酸基团，另一端即末端核苷酸的第三位碳原子上有一个游离羟基（—OH）。因此，一条多核苷酸分子就具有极性。例如两条四核苷酸 TCGA 和 AGCT（图 1－4），虽然它们的碱基组成和排列顺序相同，但它们的化学特性很不相同。习惯上把 DNA 分子序列上含有游离磷酸基的末端核苷酸写在左边，因此就把接在某个核苷酸左边的序列叫做5'方向或上游（upstream），而把

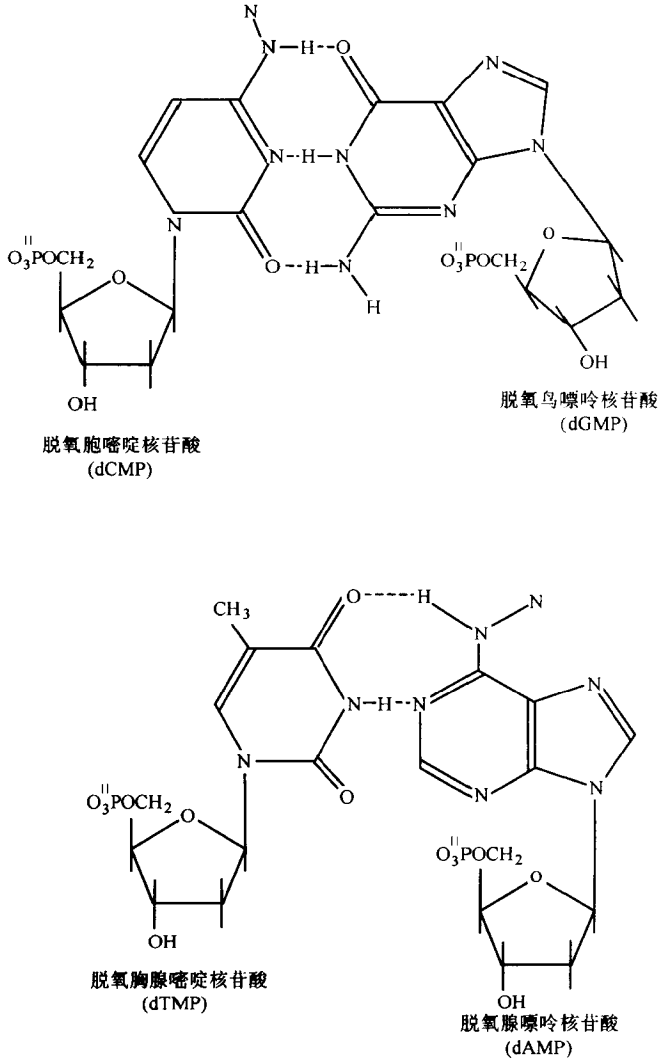


图 1－3　组成 DNA 分子的四种脱氧核糖核酸(并示氢键的形成)

接在右边的序列叫做 3’方向或下游（downstream)。

(二) DNA 分子的双螺旋结构

DNA 分子通常以双螺旋存在。沃森和克里克（J.D.Watson 和 H.F.C.Crick，1953）根据 X 射线的衍射资料、碱基的结构和 DNA 分子中鸟嘌呤与胞嘧啶以及腺嘌呤与胸腺嘧啶含量相等这三方面的信息，提出 DNA 分子是一种由两条方向相反的单链分子以右螺旋的方式，相互缠绕在一起的双螺

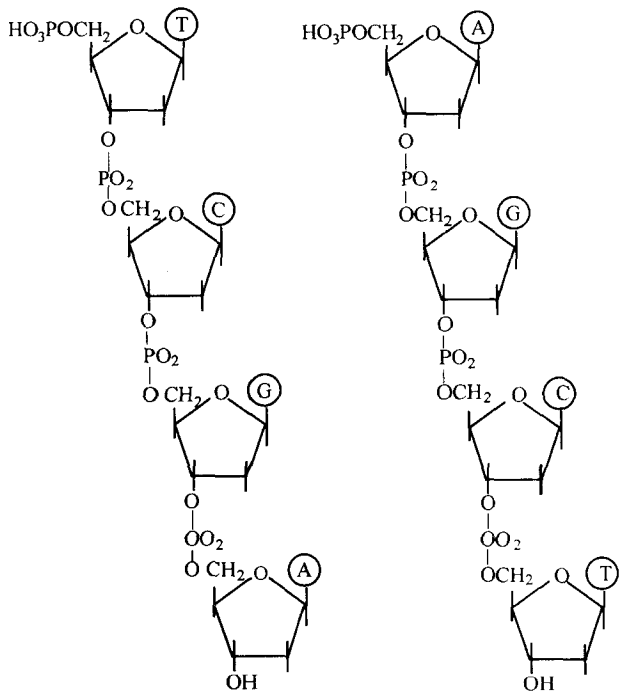


图 1－4　DNA 分子的极性

两条四核苷酸 TCGA（左）和 AGCT（右）的碱基组成和排列顺序相同但极性不同

旋（double helix）结构（图 1－5）。在双螺旋结构模型中，两条多核苷酸单链通过两链碱基之间的氢键相连，即 A 与 T 结合，G 与 C 结合，这种现象叫做碱基互补（base complementation）。A－T 和 G－C 碱基配对是固定的，虽然不同来源的 DNA 分子中 A（或 T）和 G（或 C）的含量变化很大，但是双链 DNA 分子中 G 与 C 和 A 与 T 的克分子比例始终相等，所以 DNA 分子的基本单位是碱基对（base pair，bp）。这样，所有碱基对都由一个嘌呤和一个嘧啶碱构成。在碱基对中，A 和 T 通过 2 个氢键结合在一起，而 G 和 C 则通过 3 个氢键结合在一起（图 1－5），因此 G－C 碱基对比 A－T 碱基对牢固，这就使 DNA 双螺旋的不同区域牢固程度不同。在双螺旋结构中，通过氢键连接所产生的碱基对象螺旋梯的梯级一样，与 DNA 分子的纵轴垂直，碱基对之间相距 0.34nm，双螺旋以 10 个碱基对绕纵轴一圈（图 1－5），使脱氧核糖－磷酸主链将碱基对围在双螺旋的里面。

具有极性的两条 DNA 链是反向平行的，因此，在一个 DNA 双螺旋中，只要知道一条链的碱基顺序，就可以推导出另一条链的碱基顺序。

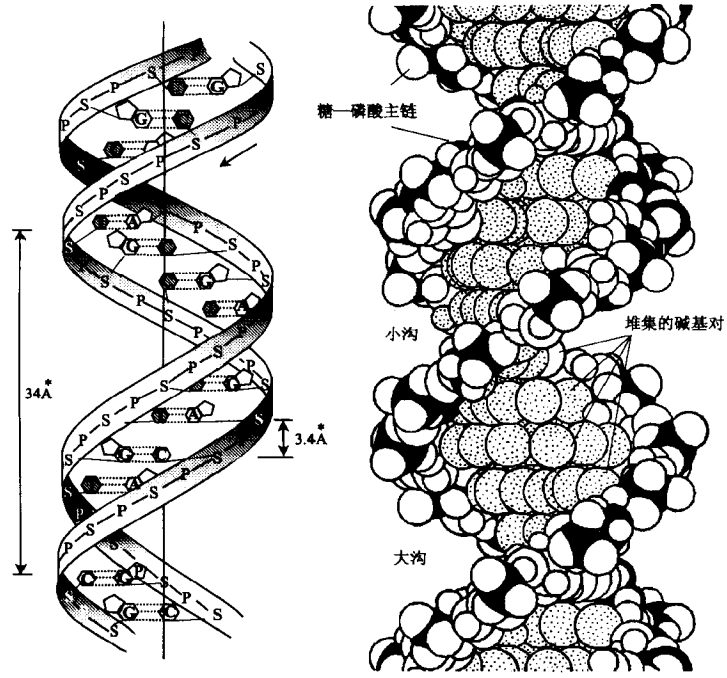


图 1－5　DNA 结构的 Watson－Crick 双螺旋模型（左）和空间一填充模型（右）

A、T、G、C 分别代表腺嘌呤、胸腺嘧啶、鸟嘌呤和胞嘧啶。S 和 P 代表糖（2－脱氧核糖）和磷酸基团

（三）DNA 的构型

DNA 分子有几种不同构型，最普遍蹬一种是 B－DNA，如上述呈右旋的双螺旋。在 B－DNA 的每一个螺旋重复上，由于两条磷酸主链相互之间稍有偏距，所以 DNA 分子中两个沟就略有不同，较大的一个叫大沟（major grove），较小的一个叫小沟（minor grove）。这种右旋的 DNA 在高浓度的盐溶液中或脱水状态下则表现为 A－DNA。B－DNA 每旋转一周约需 10.4 个碱基对，而 A－DNA螺旋更紧密，每圈具有 11 个碱基对。A－DNA 在体内不一定经常存在。在某种特定条件下，当 DNA 序列中嘌呤碱基和嘧啶碱基交替排列时，双螺旋 DNA 则以另一种构象出现，即 Z－DNA。Z－DNA 呈左旋，每圈具有 11.5 个碱基对。

DNA 分子非常长，在电镜下，10^6 个碱基对的双螺旋长约 0.34mm，直径为 2nm。在原核生物中，DNA 是环状分子，所以没有 3'－和 5'－末端。真核细胞中的 DNA 为线状分子，有3'末端和 5'末端之分。由于不同物种的染色体数目不同，所以细胞中 DNA 含量因物种不同差别很大，通常 DNA 含量愈高，

物种愈复杂。

某个细胞中的线状或环状 DNA 分子的结构一般很致密，因为双螺旋结构自身又可进一步螺旋化，形成所谓的超螺旋（supercoiling）或多次超螺旋化的结构，所以其总长度会大大降低，而其直径则相对增大。在真核细胞中，这种构型由数种蛋白质所固定。当这种 DNA 或其中一部分行使功能时，通过解螺旋，又可将双链分开成两条单链。

细菌染色体和质粒都是环状 DNA 分子，其双螺旋为 B－DNA，在形成超螺旋的过程中，DNA 双螺旋在某些区域并不相互盘绕（左向旋转），这种结构叫做负超螺旋（negative supercoiling），这是细菌 DNA 的一种重要特征。负超螺旋可以折叠起来，形成更致密的结构。如果某种 DNA 分子未超螺旋化，这种 DNA 叫做松弛型 DNA（relaxed form）。如果环状 DNA 双螺旋向右盘绕，就形成所谓的正超螺旋（positive supercoiling）。正超螺旋可转变成负超螺旋，这一反应由 DNA 回旋酶（DNA gyrase）又叫拓扑异构酶Ⅱ（topoisomerase Ⅱ）所催化。这一反应过程需要能量将 DNA 的两条链打开，以便另一部分 DNA 穿过缺口（图 1－6），最后再将缺口封闭起来。超螺旋 DNA 也可回复到松弛型，这一过程由拓扑异构酶Ⅰ催化，但不需要能量。回复过程也需要使 DNA 的一条链形成缺口，缺口形成后，断裂位点的 5'磷酸与拓扑异构酶Ⅰ上一个酪氨酰基（tyrosyl residue）结合，另一端游离的3'－OH 基团绕另一条完整链旋转一周，然后缺口再封闭起来（图 1－7）。

DNA 超螺旋还可通过双螺旋 DNA 局部解链，产生单链区而形成松弛型 DNA。单链区由各种与单链 DNA 结合的蛋白质所固定，这类蛋白质叫做单链结合蛋白（single stranded binding protein）。双螺旋打开成单链是 DNA 复制和转录的重要步骤，因为这两个过程都需要单链 DNA 作为模板。

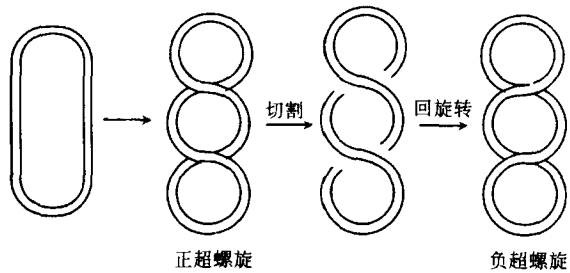


图 1－6　回旋酶与环状 DNA 分子负超螺旋的形成

（四）细胞中的 DNA 通常与蛋白质结合

原核细胞和真核细胞中的 DNA 都与某些特殊蛋白质结合，形成致密结构。真核细胞中，与 DNA 结合的蛋白质主要是组蛋白。组蛋白为一类小分子蛋白，带正电荷，富含精氨酸和赖氨酸。这种 DNA 与组蛋白结合形成的致密结构叫染色质（chromatin）。染色质的基本结构单位叫核小体（nucleosome），

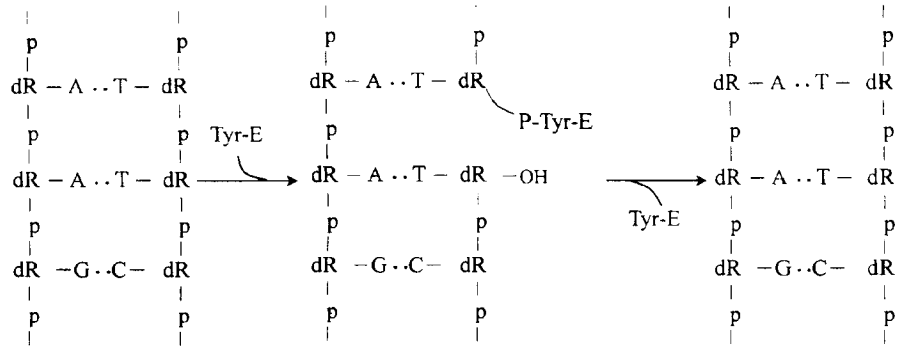


图 1－7　拓扑异构酶 I 与超螺旋 DNA 的松弛过程

Tyr－酪氨酰基　E－拓扑异构酶 I

是 DNA 双螺旋绕由四种组蛋白（H2A、H2B、H3 和 H4）形成的 8 聚体核心 2 圈而形成的一种颗粒。所以电镜下，染色质看起来就像一串念珠（第四章）。在染色质的所有蛋白质中，组蛋白约占 1/2，其余蛋白质可能与 DNA 复制和转录有关。

某些病毒 DNA 也与蛋白质结合。例如 SV40 病毒 DNA 始终都与组蛋白结合，形成长约 24 个核小体的染色质，这种结构与其寄主细胞中染色体的结构完全相同。相反，腺病毒（adenovirus）DNA 包裹到病毒粒子中后则不与组蛋白结合，当它在寄主细胞内行使功能时，释放一种由自身基因组编码的 DNA 结合蛋白，然后再与寄主细胞的组蛋白结合。

原核细胞也含有类似组蛋白的 DNA 结合蛋白。根据对大肠杆菌的研究，细菌 DNA 也与若干小分子量的碱性蛋白结合在一起，这些蛋白质都具有类似组蛋白的特征。

三、DNA 分子的功能单位

我们已经知道 DNA 分子很长，但是 DNA 分子中的功能单位平均大小却只有 1 000 个碱基对，这些功能单位相当于经典遗传学中的基因，它们决定各种表型特征的发育。基因的另一同义词叫做顺反子（cistron），但顺反子这一词使用并不很广泛。

基因含有指导合成其他化合物的信息，虽然 DNA 分子中只有四种碱基，但其可能的排列组合数却大得惊人，如果一段 DNA 含有 1 000 对碱基，那么就能形成 $4^{1\,000}$种可能的序列。由于任何生物体的遗传物质至少具有数百万个

碱基，显然，生物体就具有极其广泛的变异范围。分子遗传学中，通常就把某个生物所具有的所有基因叫做基因组（genome）。基因组中绝大多数遗传信息都用于决定蛋白质的氨基酸序列，所以基因多是通过指导蛋白质合成并通过蛋白质在细胞中执行各种重要功能。遗传信息的表达是以 RNA 为媒体进行的，但某些 RNA 本身也可直接用于各种特殊目的（本章第二节）。

基因组除了含有遗传信息的编码序列以外，在编码序列的两端还接有调节基因活性的侧翼序列（flanking sequense）。在原核生物中，当缺乏某些代谢产物时，这类基因的调控元件则决定在某个特殊阶段哪些基因将进行活动。在高等动、植物中，由于存在各种特化细胞，所以在某些器官中有些基因可能永久关闭，而另一些基因则可能正在活动。

在各种生物特别是在真核生物的基因组中，还有一些 DNA 序列没有明确功能，它们可能是基因组中新基因起源的基础，也可能是某些基因退化的遗迹。

四、DNA 分子的复制

（一）半保留复制

DNA 分子能够自我复制。当细胞分裂时，每一个子细胞都具有一整套与母细胞相同的 DNA（真核细胞的配子除外），因此 DNA 复制必定非常准确。

根据对大肠杆菌染色体 DNA 复制过程的研究，DNA 复制时，首先从一个起点将氢键打开，使双链 DNA 分开成两条单链，接着四种脱氧核糖核苷三磷酸 ATP、GTP、TTP 和 CTP 按碱基配对的原则通过氢键与每一条链上的相应碱基进行配对，然后 DNA 聚合酶（DNA polymerase）再将这些脱氧核糖核苷三磷酸连接在一起，同时释放无机焦磷酸。合成过程总是从新增长链的 3’端开始，即子链的合成按 5’→3’方向进行。在 DNA 合成过程中，两条母链并不完全分开，只是在叫做复制叉（replication fork）的位置打开，而复制叉则随着复制过程的进行不断向前移动。在新合成的两条单链中，其中有一条链可以随着复制叉不断向前移动而连续合成，这条链叫做前导链（leading strand）。而另一条新链合成时，则是首先合成一段短的单链片段，然后 DNA 连接酶（DNA ligase）再将这些新合成的单链片段连接起来，形成一条连续的单链（图 1－8），因此这条链叫做后随链（lagging strand），这些小单链片段叫做冈崎片段（Okazaki fragment）。这种 DNA 复制模型叫做半保留复制（semi－conservative replication），因为新合成的双链 DNA 分子有一条链是新合成的，另一条链则是原来的亲链。

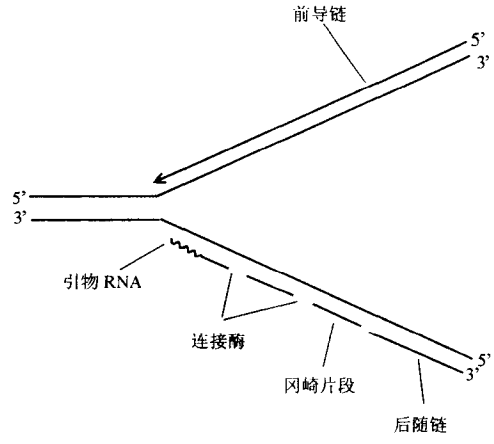


图 1－8　DNA 合成示意图

冈崎（1973）发现细菌在合成 DNA 片段之前，先由一种特殊的 RNA 聚合酶以 DNA 为模板，合成长几十个核苷酸的 RNA。这段 RNA 起引物的作用，称为引物 RNA 或引物（primer）。然后 DNA 聚合酶才起作用，以 DNA 为模板，按 5’到 3’方向合成单链 DNA 片段。引物 RNA 的 3’端与 DNA 片段的 5’端连接在一起，然后 DNA 聚合酶 I（DNA polymerase I）将引物 RNA 除掉，并在 RNA 的位置补上 DNA 片段，最后这两个相邻的 DNA 片段再由 DNA 连接酶连接在一起。这种 DNA 复制模型看起来很简单，但实际的复制过程相当复杂，涉及一系列步骤和多种复制酶，其详细过程分别见第三、四两章。

（二）复制子

在原核细胞和真核细胞中，DNA 复制时都从染色体上某个特定位点开始，这个位点叫做复制启始区（replication origin），然后复制过程再从两个方向同时进行，最后在两个复制叉会合处终止，这个位点叫做终止位点（terminus），这种复制方式叫做双向复制（bidirectional replication）。原核细胞的染色体或质粒 DNA 都是环状分子，都只有一个复制起点，所以 DNA 复制图象就像一个希腊字母 θ（图 1－9）。真核细胞具有多条染色体，每条染色体上又有多个复制起点和终点，两个相邻复制终点之间的一段 DNA 叫做一个复制子（replicon），所以真核细胞的染色体具有多个复制子，而原核细胞的染色体或质粒则只有一个复制子。

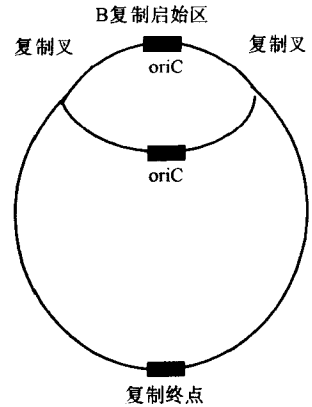


图 1－9　细菌染色体复制

从叫做 oriC 的复制启始区开始，向复制终点双向进行

对于环状 DNA 分子来说，当两个复制叉在复制终点会合后，就形成两个拷贝的环状分子。但是这两条子双螺旋并不能自动分开，而是相互缠绕在一

起，根据离体条件下的研究，拓扑异构酶Ⅱ催化这种相互缠绕的双螺旋分子的分割过程，可能活体内，环状双螺旋分子的分离也是由拓扑异构酶Ⅱ催化。

但是，对于线状双链 DNA 分子来说，其两端都有一段额外的重复序列，由于这些重复序列不是通过正常复制机器合成的，其合成过程我们将在有关章节中加以讨论。

（三）DNA 复制的滚环模型

环状 DNA 分子复制的另一种方式为滚环模型（rolling circle model）。这种复制模型适合许多病毒 DNA、细菌 F 因子以及某些基因扩增时的 DNA 复制。在这种复制模型中，当 DNA 复制开始时，环状亲本双螺旋分子首先在一条正链的复制起点上切割，然后其 5’端脱离双链分子，使 DNA 聚合酶Ⅲ以负链为模板，按碱基互补配对的原则，在正链的 3’端加上脱氧核糖核苷酸（图 1－10）。随着切割链的 5’端滚出双螺旋，游离出的单链逐渐增长，单链 DNA 结合蛋白（SSB）结合进来。由于与 SSB 蛋白结合的单链在分开时，双螺旋模板须沿其中轴滚动，所以就把这种复制结构叫做滚环。

环状双链 DNA 分子采用滚环机制复制时，正链的 5’单链部分通过迅速合成冈崎片段而转变成双链。在合成冈崎片段时，也须由引物酶首先合成引物 RNA，然后再接着合成 DNA。正链 3’端复制体的运动是由螺旋酶驱动的（图 1－10）。

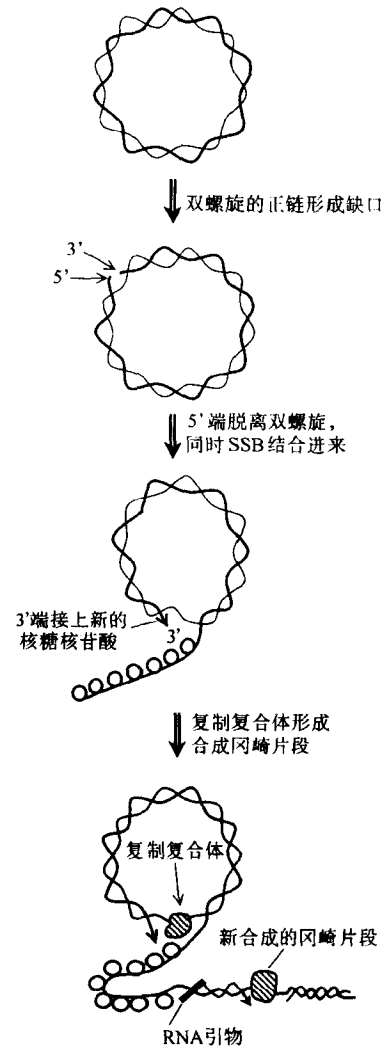


图 1－10　双链环状 DNA 分子的滚环复制模型

5’端不断解螺旋，同时螺旋酶驱动复制体移动

在这种滚环复制机制中，随着复制过程不断进行，可能产生一条含有多个基因组单位的 DNA。例如，在 λ 噬菌体 DNA 复制的后期，产生多个拷贝的基因组单位，这些多基因组单位的 DNA 分子是如何形成单个基因组的，其确切机制目前不清。一般认为，同核酸内切酶将 λ

噬菌体环状DNA分子切割成线状分子的机制相似，这一长串含有多个基因组单位的线状分子也是由某种特殊的核酸酶将其切割成一个个基因组单位。最后形成的单位基因组可能以线状分子的形式存遮，或被包裹到成熟病毒粒子中，但也有可能利用其两个单链互补末端（第八章）形成新的环状双螺旋，进行新的复制循环。

除了上述DNA复制方式以外，还有其他一些特殊的复制方式，如单链环状、单链线状病毒基因组的复制等，有关这些复制类型也将在以后有关章节中分别加以讨论。

五、DNA序列组成

如果将双链DNA分子加热或用碱处理，其A－T和G－C碱基对的氢键就可以被打开，使双螺旋结构变成单链DNA分子，这个过程就叫做变性（dissociation）。如果用高温处理，这个过程就叫熔解（melting）。一般说来，在同一温度条件下，并非所有氢键都被打断，而是在完整的双链DNA分子和完全熔解的DNA分子之间有一个过渡区，其温度相差大约是15℃，这个熔解区的中间值就叫做熔解温度或解链温度Tm（melting temperature），在0.1mol的NaCl溶液中，这个温度为95℃。

相反，如果慢慢降低温度或使pH恢复到接近中性，两条单链又可恢复到双链结构，这个过程叫做复性（reassociation）。复性速率与温度、盐离子浓度等因素有关，对大分子量的DNA来说，还与DNA片段的长度有关。如果盐离子浓度一定，那么在熔解温度以下（25℃），复性速度最高。如果控制所有这些反应条件，复性速率就与参加反应的序列的浓度和反应时间成正比。

如果令C为任一时间t秒的单链DNA摩尔浓度，Co为DNA的最初浓度，K_2为反应速率常数，复性速率就可用下列方程式表示：

$$\frac{C}{Co} = \frac{1}{1 + K_2 Cot}$$

当C/Co＝1/2时，那么$K_2 = 1/Cot1/2$，也就是说，当有一半DNA复性时，Cot 1/2与反应速率成正比。复性资料通常以Cot曲线表示（图1－11）。测定复性速率的方法有多种，既可以用分光光度计测量消色性，也可以用磷钙晶体柱将单链和双链DNA分子分开，或用S1酶处理以消除单链DNA来计算复性速率。

DNA分子的变性——复性速率是用于分析高等真核生物基因组中DNA重复情况的一种常用方法。根据复性的动力学知道，高等生物基因组中有一部分

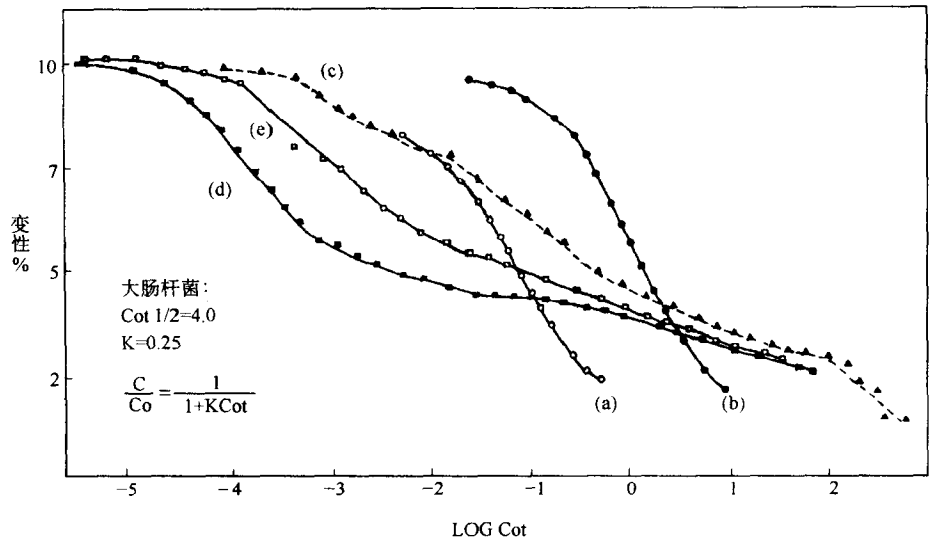


图 1－11　不同物种 DNA 变性的动力学特征

a、b、c、d、e 分别为噬菌体 T_4、大肠杆菌、黑麦 DNA、黄瓜卫星 DNA Ⅰ 和 DNA Ⅱ 的变性动态（根据 E.G.Jorden，1981）

DNA 复性非常快速，它们为一些短的 DNA 序列，其重复次数可高达 10 万个拷贝，叫做高度重复 DNA（highly repetitive DNA）。在氯化铯梯度离心后，高度重复的 DNA 常以一种染色体 DNA 主带的附属带出现，所以又叫做卫星 DNA（satelite DNA）。另一部分 DNA 在每一细胞中可重复数百倍，这类 DNA 叫做中度重复 DNA（middle repetitive DNA）。基因组中只出现一次的基因或 DNA 序列复性速度最慢，叫做单拷贝 DNA（single copy DNA）。

按重复序列的分布情况划分，高等真核生物基因组中的重复 DNA 大致可分为首尾连接（tandem array）和分散排列（dispersed array）两种类型（图 1－12）。分散排列的重复 DNA 通常分布在各个单拷贝 DNA 序列之间。在某些高等植物的基因组中，这种分散排列的重复 DNA 序列的长度都不超过 1 000bp；而对于单拷贝 DNA 序列来说，在所有植物基因组中，其长度都不超过 2 000bp，分散在整个基因组中。所以，在高等植物中，重复 DNA 的分布主要为分散排列，少数为首尾连接。

根据对 DNA 分子的变性——复性过程的研究还知道，在生物的基因组中还普遍存在颠倒重复序列（inverted repetitive sequence），其长度一般为 50～1 000bp。含有这类序列的一条单链 DNA 分子可以形成发夹环（hair－pin loop）结构（图 1－12）。具有极低 Cot 值，在电镜下也可以观察到这种结构。根据颠

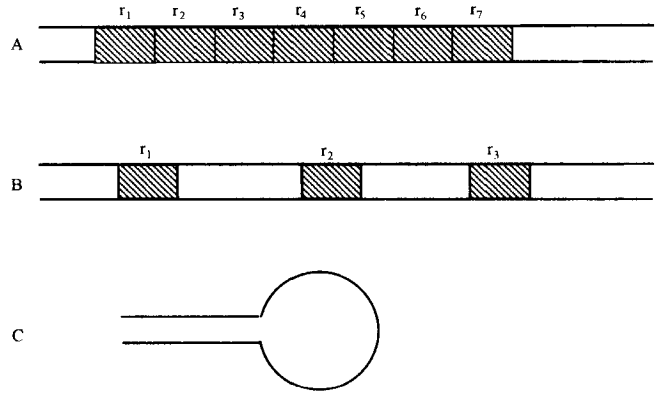


图 1－12　重复 DNA 的排列方式

A. 首尾连接的重复序列　B. 分散排列的重复序列　C. 分散排列的颠倒重复序列变性/复性后碱基配对的区域和未配对区域形成的发夹环结构

倒重复 DNA 的结构特点及其在各种生物中存在的普遍性分析，它们可能与转座子有着某种联系（第二章）。

第二节　RNA 的结构与功能

一、RNA 分子的结构

在绝大多数生物中，DNA 是遗传物质，但是在某些病毒中，RNA 也是遗传物质。在原核细胞和真核细胞中，RNA 则是 DNA 分子中遗传信息表达的媒体。

同 DNA 相似，RNA 也是由 4 种核苷酸组成，即一磷酸腺嘌呤（AMP）、一磷酸鸟嘌呤（GMP）、一磷酸胞嘧啶（CMP）和一磷酸尿嘧啶（UMP）（图 1－13）。RNA 同 DNA 一样，这 4 种核苷酸也是通过－3’和－5’磷酸二酯键连接在一起。与 DNA 不同的是，RNA 中 4 种核苷酸的糖为核糖，而在 DNA 分子中则是脱氧核糖。RNA 不同于 DNA 的另一方面是，其中含有尿嘧啶碱基而不是胸腺嘧啶碱，这两种化合物的不同点在于胸腺嘧啶中多了一个甲基。

RNA 通常以单链分子存在，由于其分子内某些区域的碱基具有互补性，即进行 A－U 和 G－C 碱基配对，所以互补碱基之间能通过氢键连接而形成发夹环和主干结构（图 1－13）。RNA 分子中，有时鸟嘌呤也可以与尿嘧啶配对，但这种碱基对不如 A－U 和 G－C 碱基对结合牢固。RNA 分子比 DNA 分

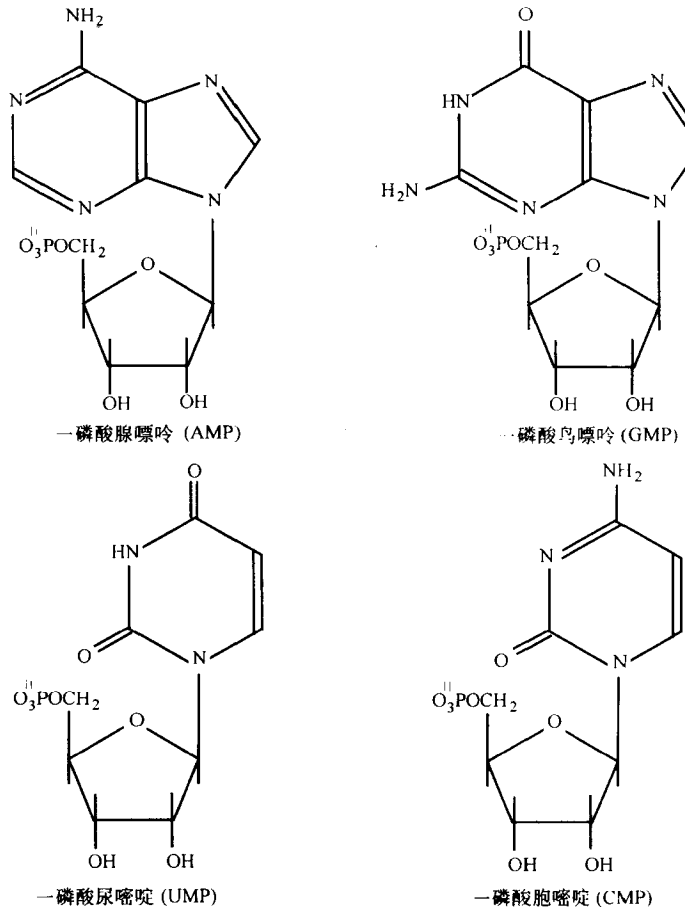


图 1－13　4 种常见的核糖核苷一磷酸

子小得多，在原核细胞和真核细胞中，DNA 分子只需用一小段来指导 RNA 分子的合成。

二、RNA 合成与病毒 RNA 遗传信息传递

（一）转录是基因表达的一个重要阶段

如前所述，基因是 DNA 分子中的一条片段，所以 DNA 分子中含有指导蛋白质合成和决定细胞内代谢过程的遗传信息。但是，DNA 并不能直接指导蛋白质合成，它必须将其所载的遗传信息复制到 RNA 上，然后再通过 RNA 来

指导蛋白质合成或控制细胞内某种代谢过程，所以 RNA 合成是基因表达的一个重要阶段。分子遗传学中就把这种以 DNA 为模板、合成 RNA 的过程叫做转录（transcription）。

在双链 DNA 分子中，只有一条链用于指导 RNA 合成，这条链叫做模板链（template strand）或有义链（sense strand），另一条与之互补的链叫做无义链（nonsense strand）。转录开始时，双链 DNA 首先在转录起点分开成两条单链，然后随着转录过程的进行，DNA 双链不断在增长的 RNA 链的前端分开，4 种核苷三磷酸（ATP、GTP、CTP 和 UTP）通过氢键与 DNA 上的互补碱基结合。这一过程所需的酶为依赖于 DNA 的 RNA 聚合酶（DNA－dependen tRNA polymerase），简称 RNA 聚合酶，它催化核苷酸之间形成磷酸二酯键，同时释放出无机磷酸，使 RNA 链按 5’→3’方向延长。大肠杆菌以及其他一些原核生物只有一种 RNA 聚合酶，而在真核细胞中，至少存在 4 种不同的 RNA 聚合酶，它们催化不同 RNA 的合成。

在大肠杆菌中，RNA 聚合酶以两种形式存在，其核心酶是由 $\alpha_2\beta\beta'$ 组成的四聚体，而全酶则含有一种叫做 σ 的额外多肽。σ 部分是准确启始转录过程所必需的，在全酶通过 DNA 模板链期间，再从全酶上脱离下来。某些情况下，RNA 合成还需要另一种叫做 ρ 的多肽以终止转录过程。有关转录的详细过程分别见第三、四两章。

（二）RNA 病毒遗传信息的传递

在各种原核生物和真核生物病毒中，有些病毒的基因组为 RNA 而不是 DNA，由于寄主细胞不能复制 RNA，所以 RNA 病毒必须提供 RNA 转录酶（RNA transcriptase）寄主细胞才能复制其 RNA 基因组。有些 RNA 病毒具有 RNA 转录酶，在侵染期间将其带进寄主细胞。有些 RNA 病毒则利用病毒基因组的遗传信息，依靠寄主细胞来合成 RNA 转录酶。

单链 RNA 病毒利用 RNA 转录酶合成一条复制型 RNA 链（replicative RNA strand），然后再以这条链为模板，合成许多拷贝的病毒基因组 RNA。复制型 RNA 链不能包裹到病毒粒子中，只用作基因组 RNA 复制的模板。

有些 RNA 病毒则首先通过合成一种 DNA 中间产物，称为互补 DNA（complementry DNA），然后再以 DNA 为模板合成基因组 RNA，这类病毒叫做逆转录病毒（retrovirus）。逆转录病毒具有一种逆转录酶（reverse transcriptase）即依赖于 RNA 的 DNA 聚合酶，该酶合成双链 DNA 分子，然后再以 DNA 分子为模板合成病毒 RNA 基因组。

三、细胞中主要的 RNA

细胞中存在许多种不同的 RNA 分子，其中最主要的有四种类型。

（一）信使 RNA

信使 RNA（messenger RNA，mRNA）载有蛋白质分子中氨基酸序列的密码信息，细胞中存在许多种不同的 mRNA，它们确定各种蛋白质的氨基酸序列，起着传递信息的作用，所以叫做信使 RNA。mRNA 通常没有次级结构。在原核细胞中，通常是几种不同的 mRNA 连在一起，相互之间由一段短的不编码蛋白质的间隔序列所隔开，这种 mRNA 叫做多顺反子 mRNA（polycistronic mRNA）。在真核细胞中，mRNA 的初级转录子为一条很大的分子，在通过剪接加工过程之后才形成成熟的 mRNA。真核细胞的成熟 mRNA 为单顺反子结构。细胞核中的大部分 mRNA 都是初级转录子或经过部分加工的转录子，叫做不均一核 RNA（heterogeneous nuclear RNA，hnRNA）。所有 mRNA 分子除了含有确定蛋白质的氨基酸编码序列（coding sequense）以外，其 5’端或 3’端还具有不编码的侧翼序列。真核生物的 mRNA 还须在转录后通过加工，将一段多聚 A 序列接到 3’末端，形成多聚 A 尾巴（poly A tail）。

mRNA 的寿命通常不很长，且容易被 RNA 酶降解，特别是在原核生物中，其寿命只有几分钟；在真核生物中，mRNA 相对说来比较稳定，其寿命短到几分钟，长到数小时甚至数天。

（二）转移 RNA

转移 RNA（transfer RNA，tRNA）为一类小分子量的 RNA（3～4S），每一条 tRNA 大约只有 80 个核苷酸。每个细胞中大约有 40～60 种不同的 tRNA，它们起着转运各种氨基酸的作用。tRNA 的转录子是一条较大的分子，其两端或中间还含有一些额外的核苷酸，在转录后的加工过程中，这些额外的核苷酸由 RNA 酶以一种特殊的方式除掉。成熟 tRNA 分子含有几种特殊碱基（图 1－14），它们是某种酶对初级转录子中 4 种正常碱基加以修饰后形成的。

各种 tRNA 都具有相似的特征，都能够形成次级结构。并且都具有 3 个碱基配对的主干和 3 到 4 个未配对的环；其二维结构为三叶草叶型，三维结构呈“L”型（图 1－15）。每一种 tRNA 都具有一段由 3 个碱基组成的，叫做反密码子（anticodon）的序列，它位于 tRNA 分子一端一个未配对的环的中央。所有 tRNA 分子的 3’末端都具有－CCA 序列，它位于反密码子相对的一端。在某种特殊的氨酰基－tRNA 合成酶（amino acyl－tRNA synthetase）的作用下，相应的氨基酸连接到－CCA 序列的腺嘌呤碱基的 2’－或 3’位上（图 1－

16)。tRNA 的 5’ 端具有碱基 G 或 C。

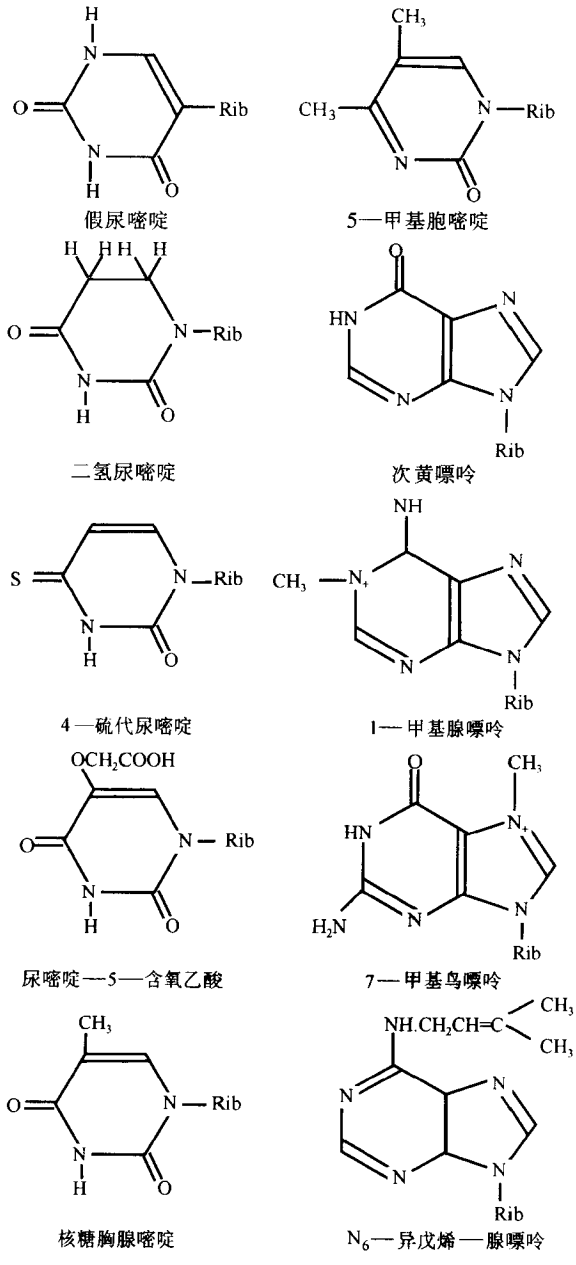


图 1－14　tRNA 中的稀有碱基

它们由 4 种最常见的碱基经化学修饰而形成　Rib－核糖

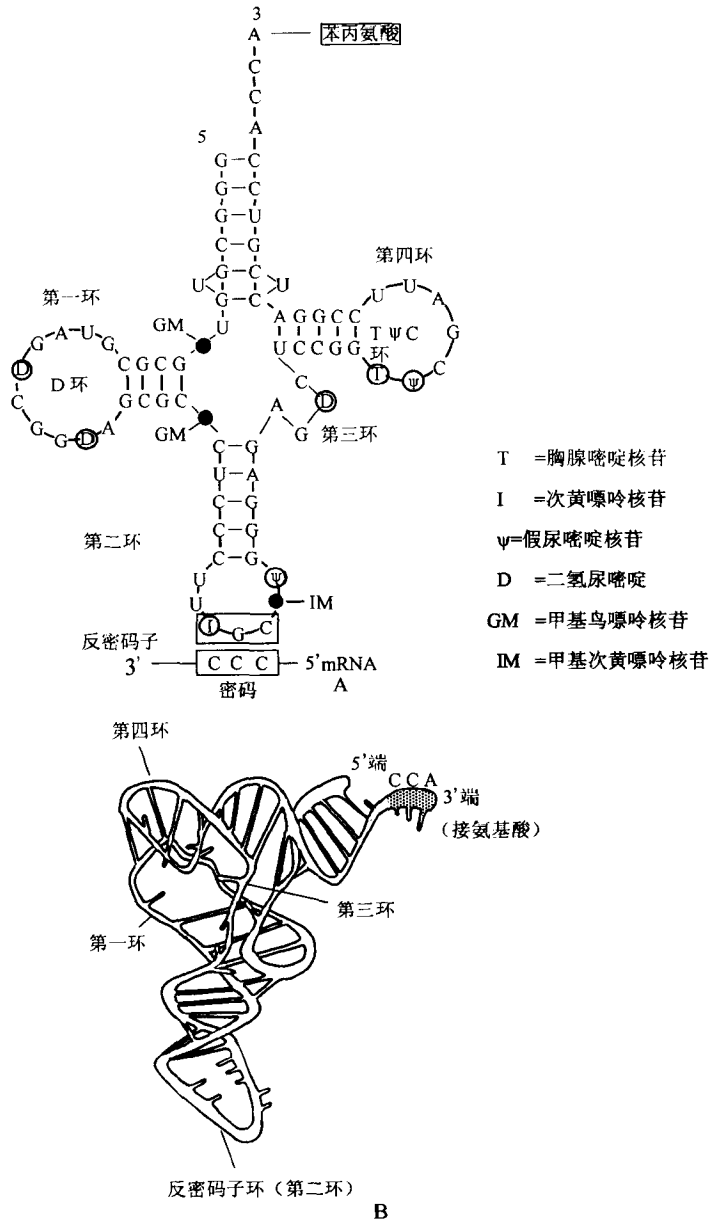


图 1－15　酵母苯丙氨酸 tRNA 构型

A. 三叶草叶型结构（二维结构）D 代表核苷酸数目不同的区域　B. 同一 tRNA 的三维结构

（三）核糖体 RNA

核糖体（ribosome）包括两个亚基即大亚基和小亚基，二者都由核糖体

RNA（ribosom－al RNA，rRNA）和蛋白质两部分组成。核糖体大亚基上有两个特殊位点，一个叫做 A 位点即氨酰基位点（aninoacyl site），另一个叫做 P 位点即肽基位点（peptidyl site），它们与多肽链合成有关。原核细胞有 3 种不同的 rRNA，而真核细胞则有 4 种 rRNA。一般认为，rRNA 的功能是在蛋白质合成过程中通过氢键把其他各种 RNA 连系在一起。这些 rRNA 分子大小不同，所以其沉降系数各不相同（表 1－1）。除了 5SrRNA 以外，其他 rRNA 都作为一个单位转录，其初级转录子是一条很大的 rRNA 分子，含有几种不同的 rRNA，并且相互之间都由间隔序列隔开。当内切核酸酶在转录后将间隔序列除掉后，才形成各种成熟 rRNA 分子。在大肠杆菌中，前体 rRNA 的间隔序列和 3’端侧翼序列还含有 tRNA 分子。无论是在原核细胞或真核细胞中，编码各种 rRNA 的基因都是重复 DNA。例如，大肠杆菌基因组含有 76 个拷贝，有些两栖动物的基因组则含有高达 500 个拷贝以上的 rRNA 基因，各个拷贝之间都由非转录的间隔序列隔开。在两栖动物卵子发生期间，当细胞对核糖体的需求急剧增加时，这些基因能够有选择地扩增，使每个细胞中 rRNA 的含量可高达 10 万个拷贝。

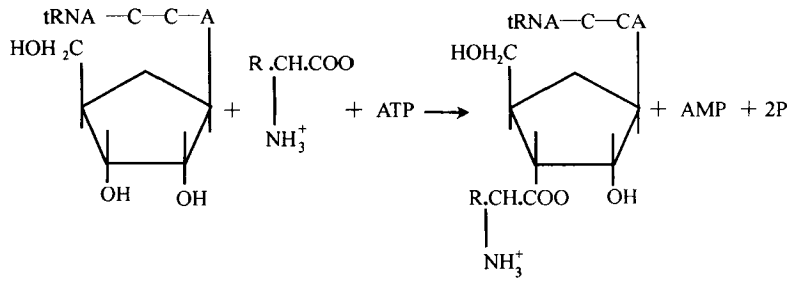


图 1－16　tRNA 分子负载相应氨基酸的过程

表 1－1　核糖体及其亚基的组成和特性

	核糖体			rRNA		
	亚基	沉降系数	结合的蛋白质种类	沉降系数	分子量×10^5	碱基数
原核生物	大亚基	50	31	23	11	2 904
				5	0.4	120
	小亚基	30	21	16	6	1 541
真核生物	大亚基	60	45	25*，26**，28+	17	4 000～5 000
				5.8	0.5	158
				5	0.4	120
	小亚基	40	33	18	7	1 800

注：*为植物，**为酵母，+为脊椎动物。

（四）核内小 RNA

核内小 RNA（small nuclear RNA，snRNA）为一类小分子量的 RNA，存在于真核生物的细胞核之中，其功能是加工其他 RNA 的初级转录子（见第四章）。由于这类 snRNA 富含尿嘧啶，所以又叫做 UsnRNA（表 1—2）。根据对酵母的研究，其细胞核中至少有 26 种 snRNA。在不同生物中，同一种 snRNA 都存在一定程度的序列同源性。各种 snRNA 通常都与大约 10 种不同蛋白质结合，形成所谓的核内小核糖核蛋白颗粒（small nuclear ribonucleo－protein particle，snRNP），参与 RNA 初级转录子的加工过程。表 1－2 中，U_4 和 U_6snRNA 通常结合在一起，其他几种 snRNA 则可以单独分离到。

表 1－2　真核细胞中部分已知 snRNA

代　号	碱基数	存在部位	功　能
U_1	165	核质	剪接外显子
U_2	189	核质	剪接外显子
U_3	210～217	核仁	不清
U_4	142～146	核质	剪接外显子、加尾
U_5	116～118	核质	剪接外显子
U_6	98～112	核质	剪接外显子
U_7	60	核质	加尾

第三节　三联体密码与蛋白质合成

一、三联体密码

DNA 分子是由 4 种核苷酸组成的多聚体，由 DNA 所编码的遗传信息在指导蛋白质合成时，以 mRNA 中三个连续排列的碱基为一组，确定蛋白质分子中一个氨基酸基团。这种由三个碱基组成的三联体叫做一个密码子（codon）。

4 种碱基三个一组，可以组成 4^3 种（64 种）不同的密码子。到目前为止，蛋白质中已知的氨基酸只有 20 种，所以在 20 种氨基酸中，有些氨基酸可以由几个密码子编码，这些密码叫做简并密码（degenerate codon）。

密码子也存在于 mRNA 上，只是 U 代替了 T。如前所述，在 tRNA 上也存在与密码子互补的三联体，这种互补三联体叫做反密码子（anticodon），密

码子上的5'碱基总是与反密码子的3'碱基配对。

经过大量实验证明，这64种密码中的大多数密码都有与之相对应的氨基酸（表1-3）。其中有3个密码子用作终止信号，不编码任何氨基酸，所以叫做终止密码（terminationcodon），有时也叫无义密码（nonsense codon）。在书写某个蛋白质基因的DNA序列时，通常以非编码链（无义链）的序列表示，因为这条链含有正确的密码顺序，所以根据其密码顺序，就可直接读出蛋白质的氨基酸顺序。

根据表1-3可以看出，当某个密码的头两个碱基确定之后，有时不管第三个碱基是什么，仍可以决定同一种氨基酸，这就是所谓的简并现象（degeneracy）。简并现象的一个重要结果就是导致同义突变（same sense mutation）。例如，当三联密码UUC突变成UUU后，后者仍然能够确定苯丙氨酸。

编码甲硫氨酸的密码AUG几乎总是用作翻译起点密码（initiation codon），所以绝大多数蛋白质的第一个氨基酸都是甲硫氨酸。当细胞中的蛋白质合成机器阅读一连串的三联密码，到达终止密码时，最后就能产生一条完整的多肽链。除了AUG用作起点密码外，还有其他一些密码如GUG和UUG也可用作起点密码，但比较少见。

tRNA在蛋白质合成过程中起着运输氨基酸与识读密码子的作用，不同的tRNA通过反密码子与密码子配对，将运来的氨基酸连接到多肽链上。但在某些情况下，tRNA中反密码子5'位的次黄嘌呤能与密码子3'位的碱基C、U和A配对，而且反密码子的5'碱基除了能与密码子第三位碱基进行正常碱基配对外，还可以进行G-U碱基配对，使一种tRNA识别一个以上的密码，因此生物体就能经济有效地利用tRNA分子。这种理论首先由Crick提出，叫做摆动假说（wobble hypothesis）。根据摆动理论，61种密码子中只有AUG和UGG各只有一个反密码子，其他密码子都有2种以上相对应的反密码子。

细胞中还存在一些同功tRNA（iso-accepting tRNA），这些tRNA具有不同的反密码子，但能运载同一种氨基酸。它们在细胞中的含量不等，含量低的tRNA在细胞中使用频率较低。另一方面，61种密码子的使用频率也不是相等的，在酵母和大肠杆菌中，编码低含量蛋白质的mRNA含有较高比例的稀有密码，例如酵母中编码3种高含量蛋白质（磷酸甘油醛脱氢酶、酒精脱氢酶和烯醇酶）的基因也只使用了61个密码子中的27个密码子。

在原生动物和真核细胞的线粒体基因组中，有些基因并不含有成熟mRNA的某些密码（图1-17），线粒体基因组在转录后，通过RNA编辑(RNAeditting)，在RNA水平上保持遗传信息的稳定性。

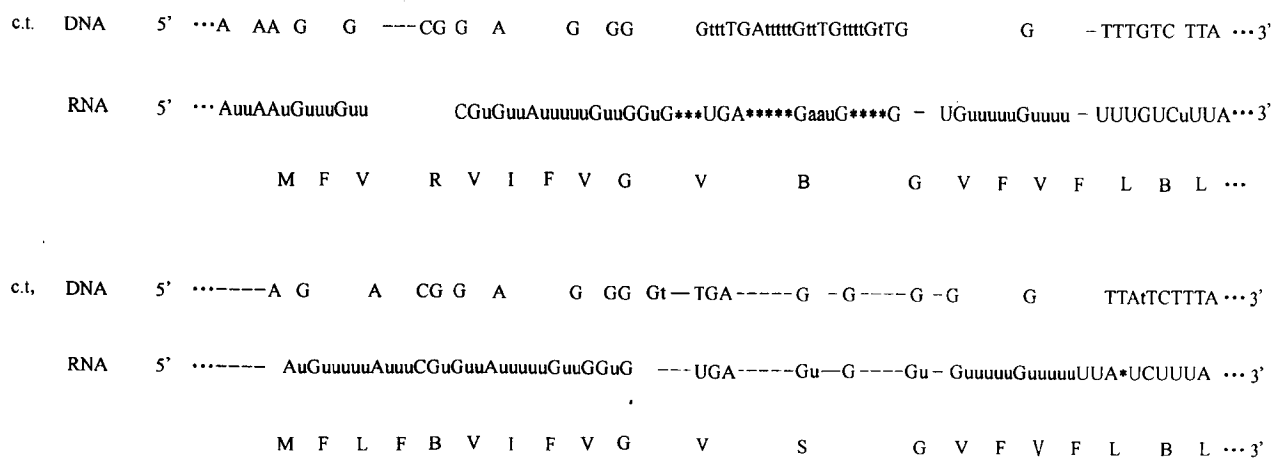


图 1-17 锥虫线粒体 CoⅢ基因的 5' 部分序列及其相对应的成熟 mRNA 的序列

U 示 mRNA 中插入的 U，* 示缺失的 T，RNA 密码下的大写英文字母示氨基酸

表 1-3　遗传密码

第一个碱基	第二个碱基 U	第二个碱基 C	第二个碱基 A	第二个碱基 G	第三个碱基
U	UUU 苯丙氨酸 UUC 苯丙氨酸	UCU 丝氨酸 UCC 丝氨酸 UCA 丝氨酸 UCG 丝氨酸	UAU 酪氨酸 UAC 酪氨酸	UGU 半胱氨酸 UGC 半胱氨酸	U C
	UUA 亮氨酸 UUG 亮氨酸（起点）		UAA 终止 UAG 终止	UGA 终止 UGG 色氨酸	A G
C	CUU 亮氨酸 CUC 亮氨酸 CUA 亮氨酸 CUG 亮氨酸	UCU 脯氨酸 CCC 脯氨酸 CCA 脯氨酸 CCG 脯氨酸	CAU 组氨酸 CAC 组氨酸	CGU 精氨酸 CGC 精氨酸 CGA 精氨酸 CGG 精氨酸	U C A G
			CAA 谷氨酸 CAG 谷氨酸		
A	AUU 异亮氨酸 AUC 异亮氨酸 AUA 异亮氨酸	ACU 苏氨酸 ACC 苏氨酸 ACA 苏氨酸 ACG 苏氨酸	AAU 天冬酰氨 AAC 天冬酰氨	AGU 丝氨酸 AGC 丝氨酸	U C
	AUG 甲硫氨酸（起点）		AAA 赖氨酸 AAG 赖氨酸	AGA 精氨酸 AGG 精氨酸	A G
G	GUU 缬氨酸 GUC 缬氨酸 GUA 缬氨酸 GUG 缬氨酸（起点）	GCU 苯丙氨酸 GCC 苯丙氨酸 GCA 苯丙氨酸 GCG 苯丙氨酸	GAU 天冬氨酸 GAC 天冬氨酸	GGU 甘氨酸 GGC 甘氨酸 GGA 甘氨酸 GGG 甘氨酸	U C
			GAA 谷氨酸 GAG 谷氨酸		A G

二、翻　　译

翻译（translation）是基因表达的最后一个阶段，即细胞中蛋白质合成机器按照 mRNA 上三联密码的顺序，将三联密码翻译成蛋白质中氨基酸序列的过程。习惯上将蛋白质的翻译分成三个连续的阶段，即翻译启始期、肽链延长期和翻译终止期。

（一）翻译启始期

几乎在所有 mRNA 的编码区中，第一个密码都是编码甲硫氨酸的 AUG，它是唯一编码甲硫氨酸的密码，但是含有能够与其互补的反密码子 CAU 的 tRNA

却有两种，其中有一种甲硫氨酸 tRNA（$tRNA_i^{fmet}$）只用于起点 AUG 密码，另一种 tRNA（$tRNA_m^{met}$）则用于将甲硫氨酸插入到多肽链的内部位置上。这种现象说明，起点密码 AUG 对启始蛋白质合成具有重要作用。在原核生物中，$tRNA_i^{fmet}$ 所负载的、用于启始过程的甲硫氨酸必须首先甲酰化，形成 N－甲酰甲硫氨酸（图 1－18）。但真核生物中用于启始过程的甲硫氨酸并不甲酰化。蛋白质在翻译后的加工过程中大多通过酶将启点甲硫氨酸切除掉，所以许多成熟蛋白质的 N－末端氨基酸并不是甲硫氨酸。

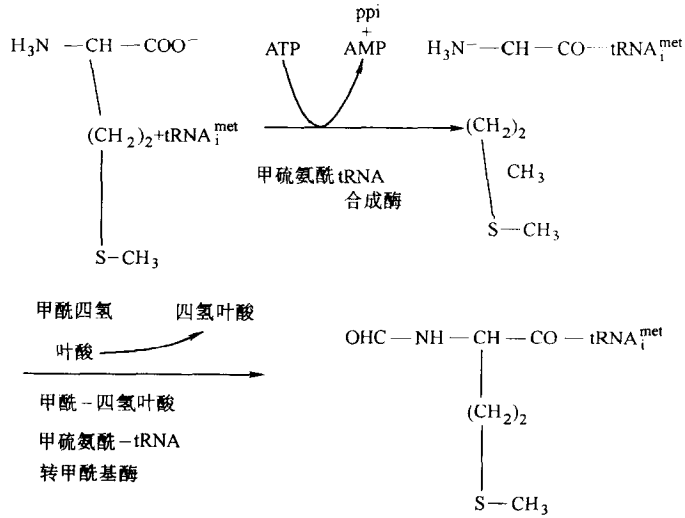


图 1－18　原核生物在启始蛋白质合成之前，$tRNA_i^{fmet}$ 的负荷过程和甲硫氨酰的甲酰化过程

在蛋白质合成的启始过程中，首先是核糖体小亚基（真核生物 40S，原核生物 30S）、fMet－$tRNA_i^{met}$ 和 mRNA 分子形成 30S 的启始复合体，然后大亚基（真核生物 60S，原核生物 50S）结合进来，形成具有功能的 70S（原核生物）或 80S（真核生物）的启始复合体。在大肠杆菌中，在蛋白质合成启始的第一步，三种启始因子（initiation factor，IF）IF1、IF2 和 IF3 附着到 30S 亚基上（图 1－19），然后 mRNA 和 fmet－$tRNA_i^{met}$ 再结合进来，其中 mRNA 通过与核糖体小亚基中 16S（真核生物为 18S）rRNA5′区域的互补序列进行碱基配对而结合在一起。这一过程还需要 GTP 提供能量，当大亚基最后结合进来后，GTP 水解，形成 GDP 和 Pi 并释放启始因子。目前，有关启始复合体形成的详细过程和一系列反应的确切顺序都不很清楚。

（二）肽链延长期

当第一个启始甲硫酸 tRNA（$tRNA^{met1}$）接到 mRNA 的起点密码上后，接

着第二个负荷的 tRNA2 就通过反密码子与 mRNA 上的下一个密码子配对，使其氨基酸基团位于核糖体的 A 位上，与起点甲硫氨酸基团相邻（图 1－20）。这一过程需要延长因子（elongation factor，EF）和 GTP。大肠杆菌中，延长因子 EF－Tu 通过与 GTP 和负载氨基酸的 tRNA 形成复合体而使负载氨基酸的 tRNA 位于核糖体 A 位上。此时，tRNA1 则已转移到核糖体的 P 位，这种肽基 tRNA 从 A 位转移到 P 位的过程需要延长因子 EF－G。然后 tRNA1 所载的甲硫氨酸基团再转移到新到的氨酰基团上，即 A 位的氨基酸上。当 tRNA1 卸下氨基酸后，脱离核糖体，将 P 位让给 tRNA2，此时 tRNA2 已载有一个二肽分子；接着第三个负荷的 tRNA(tRNA3) 又按同样方式结合到 A 位，tRNA2 上的二肽再转移到新到的氨酰基团上。按这一方式继续重复下去，于是就合成了一条多肽链。肽链延长的这种方式叫做 A 位－P 位模型，在这一过程中，核糖体阅读 mRNA 的方向始终是固定的，即总是从 5’端到 3’端，所以肽链中最先合成的一端是氨基末端，而最后合成的一端是羧基末端。

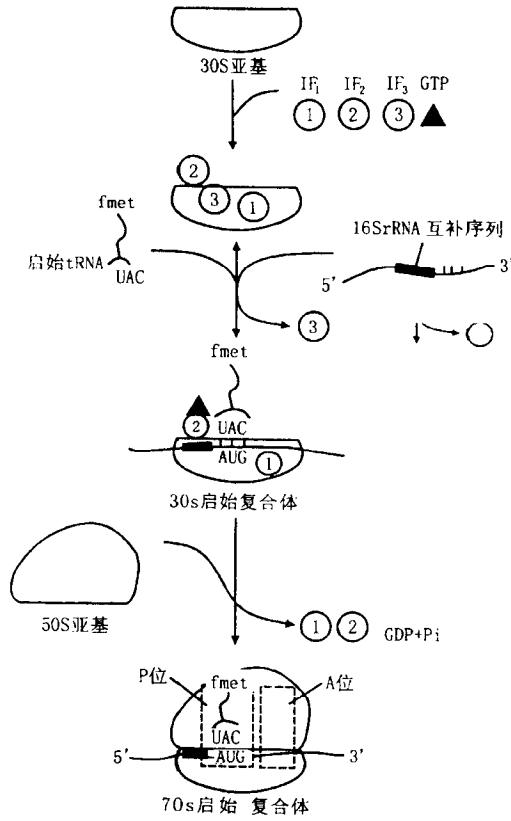


图 1－19　细菌蛋白质合成的启始过程

（三）翻译终止期

当蛋白质合成机器到达终止密码时，就进入翻译终止期。这时，所形成的多肽链从最后一个 tRNA 上水解下来。这一过程也需要几种终止因子，也叫蛋白质释放因子（protein release factor，RF）。已知大肠杆菌有三种 RF 因子，其中 RF1 识别终止密码 UAG 和 UAA，RF2 识别 UGA 和 UAA，二者的功能都是诱导肽基转移酶将增长的肽链转移到液相中。第三种因子 RF3 可能只是对这一反应起着某种促进作用。当最后一个卸下氨基酸的 tRNA 脱离复合体后，

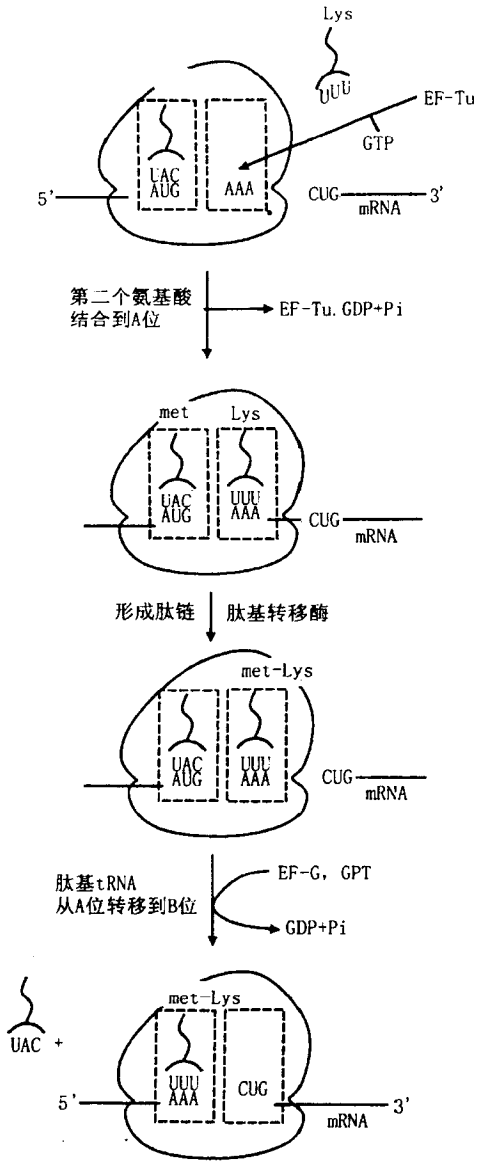


图 1－20　细菌蛋白质合成的肽链延长过程

核糖体也随之脱离 mRNA，并且其大、小亚基彼此分离，再次参与下一轮启始复合体的形成。

在电镜下常常可以见到一条 mRNA 分子上有几个核糖体，相互之间相距约 80～100nt。这种结构叫做聚核糖体（polysome）。例如编码长 145 个氨基酸的珠蛋白（globin）mRNA 大约可以容纳 5 个核糖体。所以，核糖体在蛋白质合成过程中起着装配员的作用，将氨基酸装配成多肽链，同时又起着翻译协助员的作用，因为如果 mRNA 不附着在核糖体上，遗传信息就不能翻译成蛋白质。

翻译后形成的多肽链可能还需要通过进一步加工和成熟过程才成为具有生物活性的蛋白质。如通过酶加上一个或更多的氨基酸，或者对其中某些氨基酸进行特殊修饰包括酰基化和甲基化等。在真核生物中，有些基因的产物还需要运输到内质网、高尔基体、线粒体和叶绿体中，然后再除掉多余的肽链，加上糖基或其他一些成分。有些多肽链在翻译后还要进行水解，细胞才能利用其中的功能部分。

三、中心法则及其发展

分子遗传学的一个基本原则是 DNA 将携带的遗传信息通过 DNA 分子的自我复制而延续不断地传递下去；另一方面，DNA 的遗传信息也可以转录到 mRNA 上，然后 mRNA 上的遗传信息再通过翻译，合成一定的蛋白质。这种遗传信息的流向就是遗传学的中心法则

(central dogma)（图 1－21）。

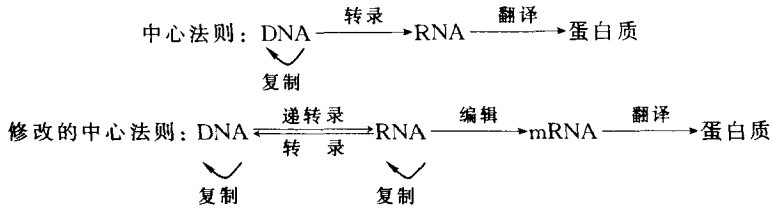


图 1－21　遗传信息的流向

后来 H.M.Temin 等（1964）发现，鸟类一种 RNA 病毒叫做劳斯氏肉瘤病毒（Rouas sarcoma virus）能够在寄主细胞中将其基因组 RNA 逆转录成 DNA，这种遗传信息由 RNA 流向 DNA 的过程叫做逆转录作用（reverse transcription）。逆转录作用丰富、发展了中心法则。

中心法则的一个基本内涵是 RNA 的基本序列正好反映其转录模板 DNA 的基本序列，但是自从 Benne 等（1986）发现 RNA 编辑以后，这一概念就受到挑战。在许多生物的线粒体和植物的叶绿体中，各种不同的、具有翻译功能的 mRNA 的编码序列都不同于其模板 DNA 的序列，这种现象说明生物体不仅以基因组 DNA 为中心，复制和传递遗传信息，而且还可在 RNA 水平上校正发生差错的遗传信息，甚至在 RNA 水平上形成新的遗传信息。

第四节　DNA 损伤及其修复

由于各种生物的生长发育都是由其自身基因所决定的，因此在生物的进化过程中，DNA 必然会发生各种变异或突变。但是进化过程中变异出现的频率是很低的，因为在每一个世代更替过程中，DNA 分子都要进行复制，以保持其遗传稳定性。对于一个生活的细胞来说，其生命过程需要数千种具有正常功能的蛋白质，而编码每一种蛋白质的基因则可能在许多位点上发生突变，从而影响蛋白质的正常功能。DNA 分子不仅会自发地出现差错如个别碱基的缺失和插入等，而且还可能受到许多化学因子或物理因素的作用而使其糖－磷酸骨架折断或使其碱基的化学特性发生改变。当 DNA 出现损伤时，其碱基的编码特性就会发生改变，如果细胞中没有某种特殊的修复机制来修复受损伤的位点，那么细胞就不能生存。自然界的各种生物都能通过 DNA 修复以降低自发突变和诱发突变的水平，使其整体遗传变异与原有遗传信息的稳定性之间保持平衡。根据对大肠杆菌和鼠沙门氏菌 DNA 修复机制的遗传和生化研究表明，DNA 修复是一种复杂而又多样化的过程。本节主要讨论细菌中各种 DNA 结构变异的修复机制以及各种修复机制之间的相互关系。

一、DNA 变异

DNA 分子的各种组成元件对一系列损害因子都很敏感。这些损害因子或打断磷酸二酯键，或切割 N－糖苷键，或改变嘌呤碱和嘧啶碱的结构，使不同 DNA 链共价交叉连接。当然，DNA 分子也能自发产生各种结构变异。

（一）紫外光诱导的光产物

太阳光中的紫外线（UV）能诱发 DNA 出现各种结构损坏，在这些紫外光损坏产物中，研究最为详细的是嘧啶二聚体（pyrimidine dimer）。这类二聚体是通过两个相邻嘧啶碱基之间的 5’－5’和 6’－6’双键结合形成的，其中最常见的是 TT 二聚体，但也能形成 CT 二聚体和 CC 二聚体（图 1－22）。

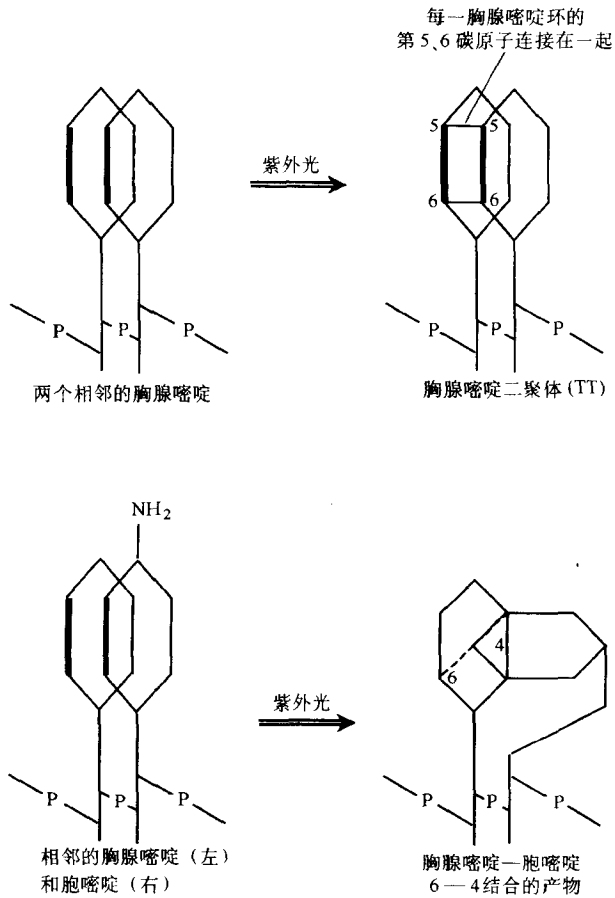


图 1－22　紫外光诱导的两种重要的嘧啶二聚体

例如，某些人患有着色性干皮病（xeroderma pigmentosum XP），该病是一种DNA修复酶缺乏病，当这种病人的皮肤暴露在正常日光下时，一天内每个细胞中的DNA能形成数千个嘧啶二聚体，由于这种病人的细胞缺少除掉这类二聚体以及其他由紫外光诱导的损害斑的DNA修复酶，因此阳光中的紫外线会杀死大量皮肤细胞，并极大地增加这种病人患皮肤癌的危险性。

根据对紫外光诱发突变的大量研究，紫外光除了诱发嘧啶二聚体形成以外，还可能诱发其他各种DNA结构变异，包括DNA单链和双链断裂等结构损坏。

（二）交叉连接

有些具有双重功能的碱性诱变剂如补骨脂素（psoralen）和丝裂霉素C（mytomycinC）等都能引起DNA链之间发生交叉连接（cross link）。如果是丝裂霉素C，可造成两条链的嘌呤碱基交叉连接；如果为补骨脂素，则造成两条链的嘧啶碱交叉连接。

（三）烷化产物

许多烷化剂如甲基甲烷磺酸盐（methyl methane sulfonate）、乙基甲烷磺酸盐（ethyl methane sulfonate）等能将其甲基或乙基转移到嘧啶环、嘌呤环的各个位置以及磷酸二酯键上。烷化作用的位置主要是鸟嘌呤的N_7和C_6位的氧原子、腺嘌呤的N_3位。此外这些烷化剂也能在DNA的任何氢、氧、磷原子上进行烷化作用，但频率较低。

（四）脱氨基产物

在大肠杆菌和鼠沙门氏菌中，DNA分子中的胞嘧啶能以较高频率进行脱氨基作用，形成尿嘧啶。如果对DNA分子的这种结构变异不进行修复，则造成G－C到A－T碱基对的转换。5－甲基胞嘧啶脱氨基后，可形成胸腺嘧啶，腺嘌呤脱氨基后，可形成次黄嘌呤，因此也能造成转换。

（五）脱嘌呤和脱嘧啶作用

一般说来，嘌呤或嘧啶碱基的甲基化作用往往破坏N－糖苷键。如鸟嘌呤甲基化后可形成7－甲基鸟嘌呤，其结果就会使DNA分子立即发生脱嘌呤作用，形成无嘌呤位点。同样，如果DNA分子中某个嘧啶碱基甲基化，就会发生脱嘧啶作用，形成无嘧啶位点。这些无嘧啶、无嘌呤位点统称为AP位点（apurinic and apyrimidinic site）。此外，细胞内的各种糖基化酶（N－glycosylase）也能产生AP位点。

（六）受损坏的核苷酸

化学诱变或电离辐射都能破坏核苷酸的氢键，形成一些氢键特征发生了变化的化合物如羟氨（hydroxylamine）和过氧化氢等。羟氨特别能破坏胞嘧啶，

而过氧化氢则能切割磷酸二酯键和破坏各种碱基。

二、DNA 损伤的修复机制

（一）光修复

光修复是指细胞内的 DNA 受到紫外光损伤后，在正常可见光下，光诱导系统对损伤位点进行修复的过程，所以这一修复过程又叫光复活（photoreactivation）。现在知道，大肠杆菌中有两种光复活酶（photoreactivation enzyme），这两种酶能切割嘧啶二聚体，将二者分开。

（二）切割修复

切割修复（excision repair）有几种途径，其中包括除掉受损的碱基和 DNA 分子的糖磷酸骨架片段。各种切割修复途径都产生相同的中间产物，或者是一段单链缺口、或者是 DNA 分子中的一个小缺刻，从而为 DNA 聚合酶提供一个 3’羟基末端，以启始受损坏区域的修复合成过程。另一种修复酶为 DNA 连接酶，它使新合成的修复片段两端形成磷酸二酯键。所以，当 DNA 分子丢失其中一个单链片段后，其必需的遗传信息可以从另一条互补单链中得到补充。

大肠杆菌中，切割修复涉及三种重要的蛋白质，即 UvrA、B、C，三者的分子量分别为 114kd、84kd 和 70kd，它们都是启始切割修复的内切核酸酶的亚基。切割修复过程大抵可以分成三个不同步骤（图 1－23）。首先，UvrA 蛋白随机地结合到 DNA 分子上，并形成 UvrA 二聚体蛋白。这一过程需要 ATP。UvrB 蛋白并不直接与 DNA 结合，而是结合到 DNA－UvrA 复合体上，提高 UvrA 蛋白对 DNA 损伤位点的特异性，形成稳定的 UvrA－UvrB－DNA 复合体。然后，UvrC 蛋白结合到这种复合体上，切割受损坏位点两侧的磷酸二酯键，除掉 12 个碱基的片段。接着，DNA 螺旋酶Ⅱ（DNA helicaseⅡ）与 DNA 双螺旋结合，释放 UvrABC 切割酶。最后，DNA 分子上长 12 个碱基的缺口由 DNA 聚合酶Ⅰ和 DNA 连接酶修复。

UvrABC 酶能作用于不同类型的损坏斑，但该酶并不直接作用于任何特殊损伤斑，只是识别正常 DNA 的缺失位点，它在双螺旋的同一侧、约相距一个螺距的两个位点切割含有损伤斑的一条单链。

（三）重组修复

切割修复使用互补链作为模板，合成一条新的单链片段以替换损坏的 DNA 片段。但有时这种模板并不能随意得到。例如，当复制叉遇到损坏斑（如嘧啶二聚体）时，或在切割修复之前，双链 DNA 就已分开成单链时，一

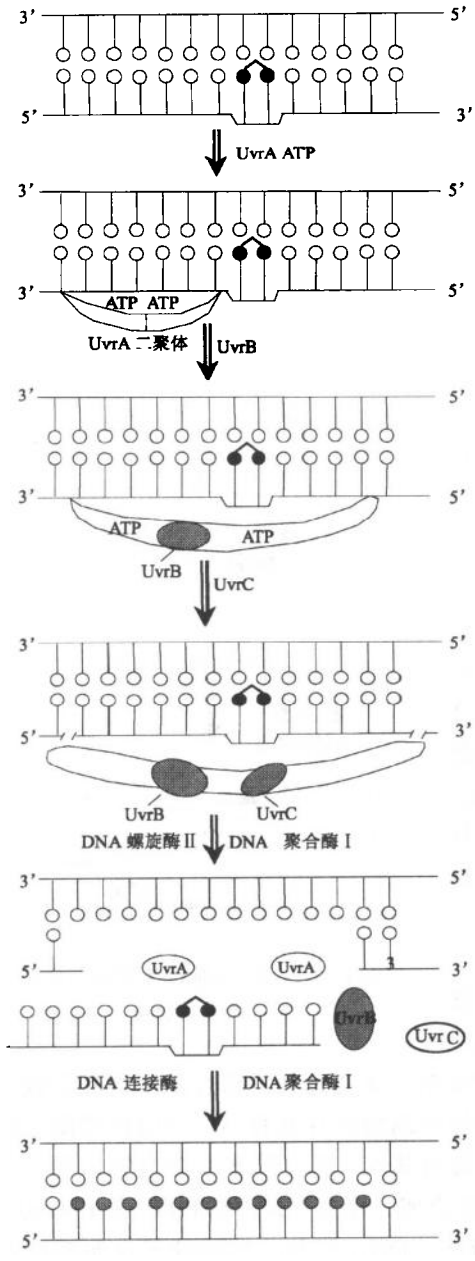


图 1－23　UvrABC 切割酶的切割修复过程

黑点为嘧啶二聚体。由 DNA 聚合酶 I 进行修复合成的片段超过切割片段的长度(阴影核苷酸)

条单链中的损坏位点就不能得到修复。另外，如果互补碱基对的两个碱基都发生变化，如由癌诱变剂丝裂霉素 C 引起的交叉连接，则两条链都不能作为对方的模板链。还有，当双螺旋齐头断裂，且双链片段完全丢失时，这种双螺旋就不能直接进行修复。在上述所有情况下，受损坏位点的遗传信息都丢失了，它们都只能从另一条具有相同序列的 DNA 分子上获得相应的 DNA 片段，否则就不能恢复到原状。这种修复机制叫做重组修复（recombinational repair）。细胞中几乎始终存在与受损位点相对应的 DNA。例如当一个复制叉通过某条单链片段的损坏斑时，第二条 DNA 的子链就含有相同的序列。或者，如果两条单链都在未复制的区域被损坏，由于生长迅速的细菌细胞往往含有一个拷贝以上的染色体，因此这些染色体都可用作修复的模板。高等真核生物通常都是二倍体，在每一对同源染色体上都存在相同或近乎相同的 DNA 序列。但是，重组修复的关键是需要一种酶将损坏斑一侧的序列并列到未损坏的 DNA 的相应序列上，使丢失了遗传信息的损坏部位与正常 DNA 的相应区域对应在一起。在大肠杆菌中，RecA 蛋白就具有这种功能。

通常认为，交换是创造遗传变异的一种机制，可能交换的最重要的功能是修复损坏的 DNA。在修复过程中，RecA 蛋白首先与缺口结合，然后再诱导另一条同源双螺旋的一条单链

填补到缺口上，使缺口得到修复（图 1－24）。虽然 DNA 聚合酶能够修补简单的缺口，但对一些结构复杂的缺口如含有嘧啶二聚体的缺口来说，当复制叉遇到嘧啶二聚体时，由于不能进行碱基配对，复制机器必须越过受损坏的位点，重新启始复制过程，因此使该位点上两条单链上的遗传信息都丢失了。这些丢失的遗传信息只能通过与同源双螺旋进行重组而获得。

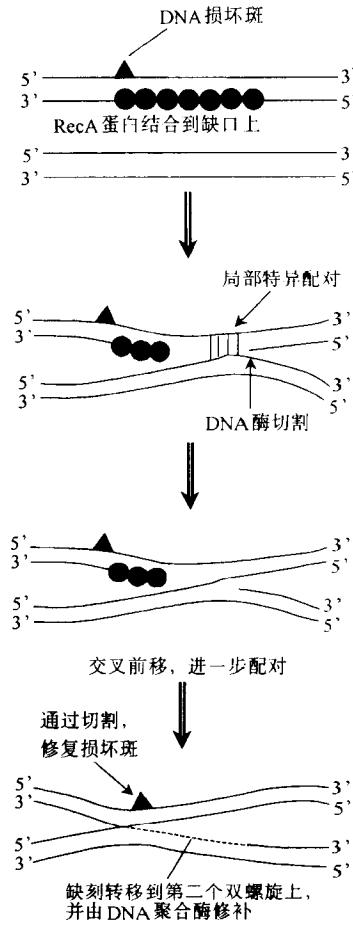


图 1－24　RecA 蛋白与缺刻结合，启始重组修复

RecA 蛋白还能修复双链断裂的 DNA 分子。两个断裂末端先由外切核酸酶降解，产生单链末端，然后两个单链末端再分别侵入另一双螺旋的同源序列之中，形成两个连接点（图 1－25）。另外，在启始修复合成之前，可能还需要另一种酶即 RecBCD 酶，这种酶通过与每一个游离的单链末端结合，为修复合成提供启始位点，然后再由 DNA 聚合酶修补缺口。当两个连接点被切割之后，两条亲本双螺旋分开，最后形成的缺刻再由 DNA 聚合酶修补和由 DNA 连接酶缝合。

（四）糖基化酶修复

大肠杆菌还有第二种切割损坏位点的方式，即 DNA 糖基化酶（DNA glycosylase）切割途径。这种酶能识别单个异常碱基，并通过切割糖苷键，将其从脱氧核糖上除掉，产生前述 AP 位点。大肠杆菌细胞含有几种 AP 内切核酸酶，这些酶既能切割 AP 位点的 3’端，也能切割其 5’端，留下一个引物末端。最后 DNA 聚合酶再从末端开始，启始修复合成，补上丢失的核苷酸（图 1－26）。

细胞中存在多种糖基化酶。如尿嘧啶－DNA 糖基化酶能校正由胞嘧啶自发脱氨基作用引起的问题。DNA 分子中胞嘧啶能自发脱氨基，形成尿嘧啶，这种糖基化酶则能识别 DNA 分子中的尿嘧啶并将其除掉，最后由 DNA 聚合酶修补缺刻。又如，腺嘌呤通过脱氨基作用可形成次黄嘌呤。现在知道，大肠

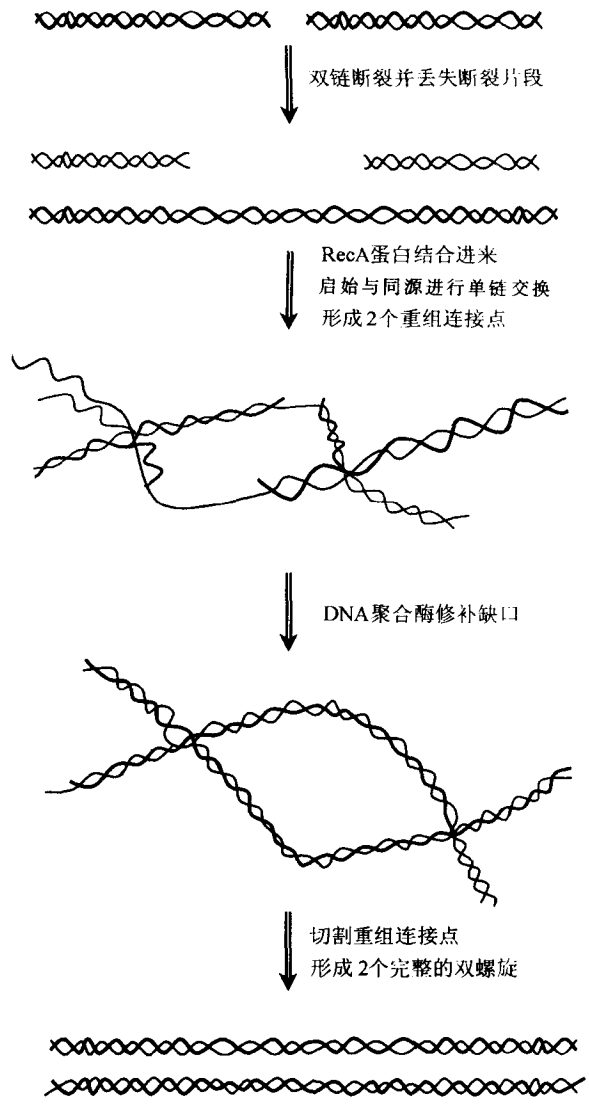


图 1－25　RecA 蛋白对 DNA 双链断裂的修复模型

杆菌中有一种次黄嘌呤糖基化酶，该酶能按上述尿嘧啶修复途径，使 DNA 聚合酶在脱嘌呤位点进行修复。

（五）误配对修复

DNA 在复制和在同源重组过程中常产生一些错误配对的碱基对，据估计，DNA 聚合酶每复制 10^8 个碱基对就会出现一个碱基的差错。在大肠杆菌和鼠沙

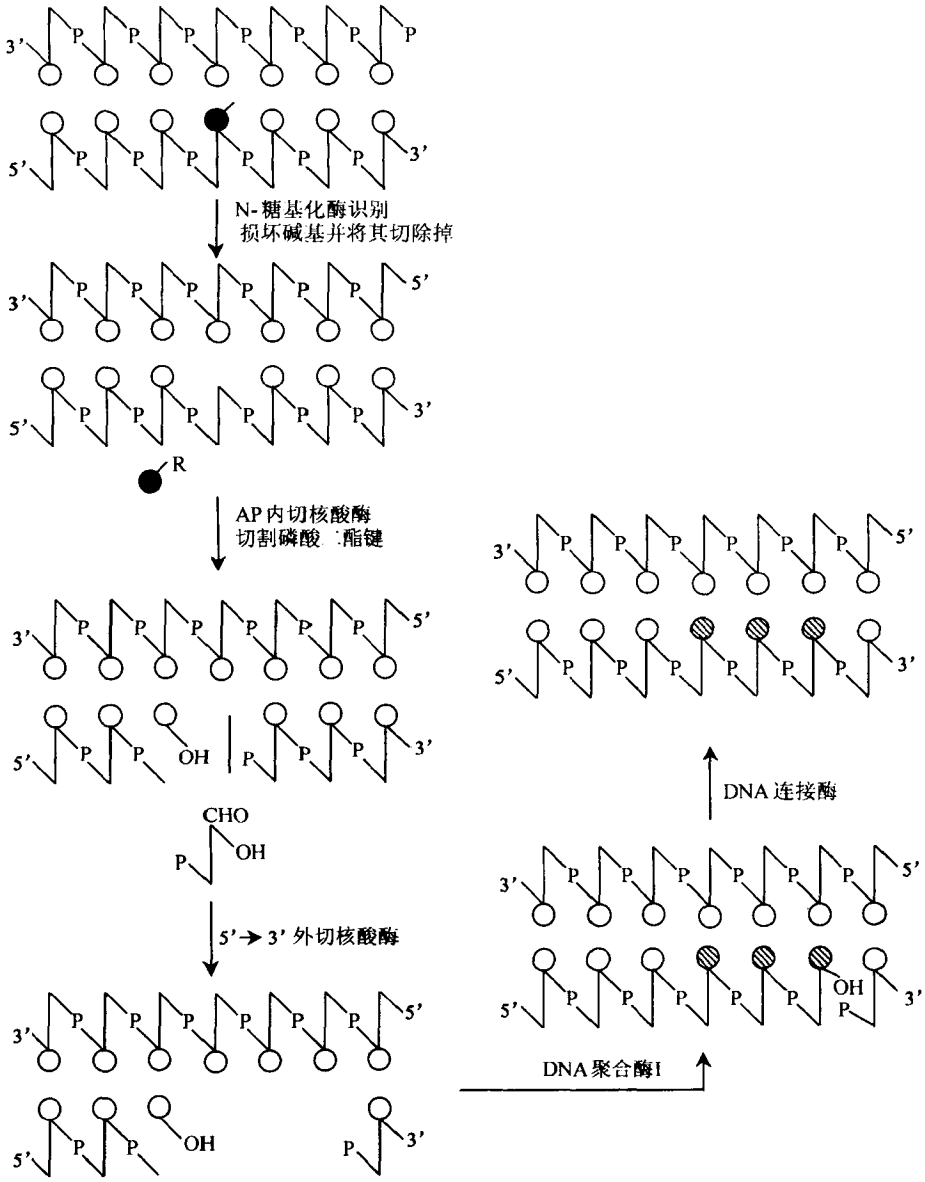


图 1－26　N－糖基化酶修复机制

R 为氨基；AP 内切核酸酶切割 AP 位点 5' 端（也可切割 3' 分端）；5' →3' 外切核酸酶最后去掉糖和磷酸；涂阴影的为修复合成的核苷酸

门氏菌中都存在误配对修正酶（mismatch correction enzyme），这种酶由三个基因即 mutH、mutL 和 mutS 编码，它能在新复制的 DNA 上寻找误配对的碱基对，并将含有错误核苷酸的单链片段除掉，使 DNA 聚合酶在修补缺口时插入

正确的核苷酸。显然，这一修复过程的关键是确定误配对的碱基中哪一个是错误的碱基，因为两个碱基都是 DNA 分子的自然组成成分。修正酶能识别含有错误碱基的一条链中 GATC 序列及其周围的序列，并通过与这段序列以及错误碱基所在位点结合，将含有误配对碱基的一条单链片段除掉（图 1－27）。

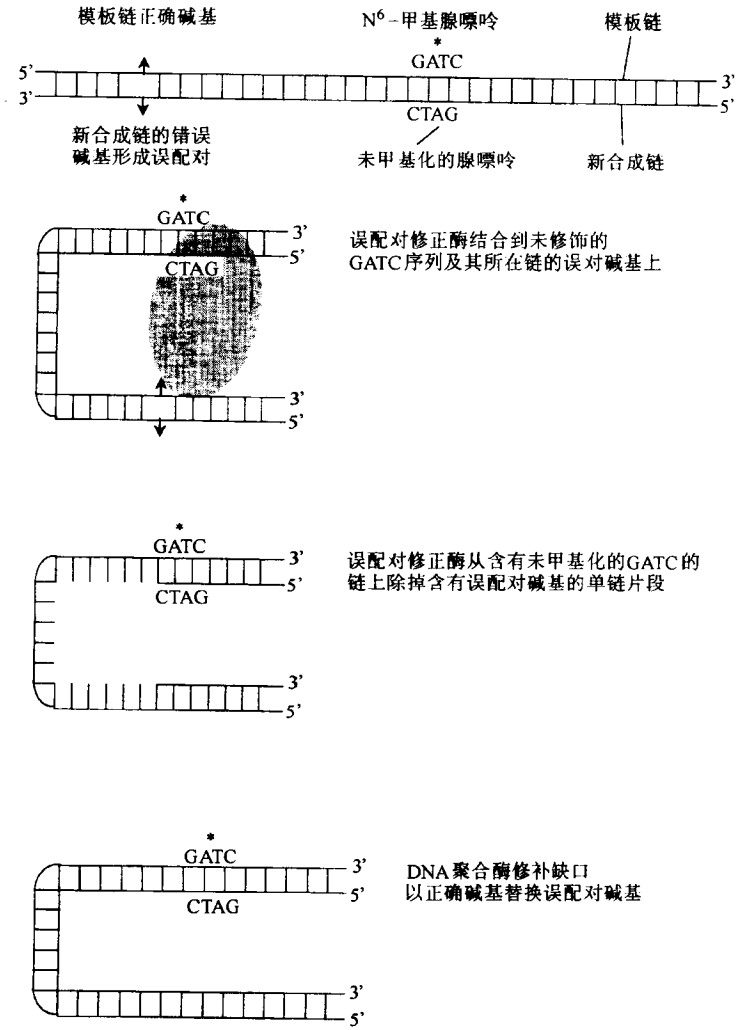


图 1－27　大肠杆菌误配对修复模型

GATC 序列是由 dam 基因编码的甲基化酶作用的位点，如果 DNA 分子在合成之后，甲基化酶首先将 GATC 序列中的腺嘌呤修饰成 N_6－甲基腺嘌呤，那么修正酶就不能进行切割。大肠杆菌中大多数 GATC 在合成之后都受到甲基化

酶的修饰，但其修饰过程往往会延迟数秒钟或几分钟，这段时间就足以使误配对修正酶与其结合，所以修正酶只与新合成链结合而不与模板链结合。误配对修复酶不仅只是确定误配对碱基，而且还能识别小片段插入和缺失区，所以它具有降低移码突变和碱基替换的作用（第二章）。

（六）倾向错误修复

某些碱基类似物和某些修饰过的碱基可通过不正常配对而引起突变，但现在还不清楚为什么损坏的碱基不配对时会引起突变。大多数自然诱变剂以及许多致癌物质都能引起突变，尽管在损坏位点上完全缺少模板信息，但细胞仍然能够进行 DNA 复制，如由紫外光诱导的 6－4 嘧啶二聚体以及 AP 位点等。细菌细胞能够修复具有 AP 位点的 DNA 分子，但在修复过程中往往造成突变，因为该位点不具有用于修复单链 DNA 分子的正确碱基，所以这种修复过程叫做倾向错误修复（eror－prone repair）。倾向错误修复可能是细胞赖以生存的最后一招，它只有在既无 DNA 模板，但又须准确进行修复时才使用的一种修复方式。因为细胞吸收某种错误碱基以保证复制继续进行比完全停止 DNA 复制要好得多。倾向错误修复造成突变的频率很高，因此这种修复途径是造成突变的主要原因之一。

大肠杆菌中，与倾向错误修复有关的两个基因为 umuC 和 umuD，二者的编码产物阻止紫外辐射诱发 DNA 损坏。这两个基因属于 SOS 基因，当用某种 SOS 诱导物处理细胞后，其编码产物大量合成，所以倾向错误修复有时也叫 SOS 修复（SOS repair）。这两个基因中任何一个发生突变都会使细胞极易于遭受紫外光杀伤而不能存活，因此这两个基因的作用是通过修复损坏位点，促进细胞生存。

大多数突变都只是在某个损坏位点发生单个碱基的变化，而 UmuC 和 UmuD 蛋白只是将单个碱基插入到增长链中，这一过程现在叫做附属合成（bypass synthesis）。在这个过程中，UmuC 和 UmuD 蛋白可能通过抑制 DNA 聚合酶 I 的 3’→5’ 外切核酸酶活性的校对作用。另一种观点则认为，UmuC 和 UmuD 蛋白完全是一种新的 DNA 聚合酶，这种酶能造成 DNA 复制错误。倾向错误修复过程还需要 RecA 蛋白，由于 RecA 蛋白与 DNA 分子中单链缺口结合，它可能通过与 UmuC 和 UmuD 蛋白结合，促进这两种蛋白质结合到需要进行倾向错误修复的位点上。

（七）真核细胞的修复系统

有关真核细胞中 DNA 损坏的修复机制目前知道得不多。在酵母中已分离到与修复受辐射损坏有关的基因，这些基因统称为 rad 基因。rad 基因可以分成三组，第一组与切割修复有关，第二组与复制后修复有关，第三组基因与重

组修复有关。各种修复类型都与转录过程有关，都倾向性地修复具转录活性的基因，而且是首先修复模板链。可能修复机器与 RNA 聚合酶有某种联系。现在知道，rad 基因之一的 rad3 基因编码一种螺旋酶（helicase），它是与 RNA 聚合酶结合的一种转录因子，是对损坏斑的切割步骤所必需的。

在哺乳动物中，由紫外线辐射或相当因素引起的 DNA 损坏斑的修复同原核细胞的切割修复机制相似。根据对人体着色性干皮病皮肤细胞的研究，XP 病人的成纤维细胞不能切割嘧啶二聚体。在人类，这种疾病大约有 9 个遗传互补群，所有病例都是在修复过程的切割步骤有缺陷。已知 XP 座位上的基因都与酵母的 rad 基因存在同源性，表明切割修复机制在各种生物中是高度保守的。

现在还发现了在哺乳动物中存在重组修复机制，这些修复系统都与遗传重组过程有关。真核细胞中还可能存在其他各种修复系统，与原核生物相比，修复机制可能更为复杂。

小　　结

两个经典实验证明 DNA 是遗传物质。第一个实验是从肺炎双球菌的一个菌株中分离到的 DNA 可将该菌株的遗传特性提供给另一菌株。第二个实验证明只有 DNA 才能从亲代噬菌体传递给子代。现在已有多种方法用 DNA 转化真核细胞，使真核细胞获得新的遗传特性。

DNA 是一种由 2 条反向平行的单链组成的双螺旋结构。每条单链都是通过两个相邻核苷酸之间形成 5’－3’磷酸二酯键连接起来的高分子化合物。两条单链再通过形成 A－T、C－G 碱基对，形成双螺旋结构。双螺旋结构在复制起始区分开成两条单链，根据碱基互补配对的原则，以半保留复制方式合成新的子链。

在某些病毒中 RNA 是遗传物质，但在真核细胞和原核细胞中，RNA 是以 DNA 为模板转录而成的。细胞中的 RNA 主要有 4 种，即 mRNA、tRNA、rRNA 和 snRNA，其中 snRNA 为真核细胞所特有。mRNA 载有 DNA 分子中的遗传信息，其他几种 RNA 都与遗传信息的表达有关。

DNA 分子以其一条单链的三个碱基组成一个密码，决定蛋白质分子中一个氨基酸。在 64 个密码中，有 61 个编码氨基酸，3 个为翻译终止信号。DNA 的两条链中只有一条链编码蛋白质，它将其遗传信息转录到 mRNA 上，然后以 mRNA 中密码子的顺序决定蛋白质分子中氨基酸的顺序。

DNA 会因各种内、外因素的作用引发结构变异或损坏，但细胞中有各种

修复机制修复受损坏的区域。细菌中有一种光复合酶消除嘧啶二聚体。切割修复是在受损坏位点附近切割 DNA，除掉一条单链并合成一段新的序列以替换被切割的片段。重组修复是从任何一条具有相同序列的 DNA 双螺旋中获取信息，并利用这种信息来修复双链都受到损坏的区域。细菌的 RecA 蛋白能够将两条同源 DNA 双链并列到一起，启始重组修复。糖基化酶修复是通过糖基化酶切割异常碱基和通过 AP 内切核酸酶切割 AP 位点，然后合成一段新的序列以替换损坏斑。误配对修复是通过 Dam 甲基化酶除掉含有错误碱基的一段序列，然后通过修复合成一段互补序列。当细胞中不存在 DNA 模板时，细胞采用任何碱基来修复损坏斑，其结果会造成基因突变，这种修复机制叫做倾向错误修复或 SOS 修复。

复　习　题

1．“转化”一词首先由格里菲斯提出，它的内涵是什么？

2．DNA 分子的特有碱基为------，RNA 分子的特有碱基为-------。

3．DNA 分子 5’端的游离化学基团为------，3’端的游离化学基团为-------。

4．能与下列单链 DNA 分子互补的 DNA 链的碱基序列是什么？

A．5’AGTTACCTGATCGTA3’

B．5’TTCTCAAGAATTCCA3’

5．DNA 分子的 A－T 和 C－G 碱基对中，哪一种碱基对较易被打开？为什么？

6．在一种特殊 DNA 分子中，$\frac{A+T}{G+C}=1$，这种 DNA 是双链还是单链？

7．一段双链 DNA 分子中的胞嘧啶核苷酸为 17%，这段 DNA 中腺嘌呤核苷酸的百分率是多少？

8．某噬菌体 DNA 含有 33%A、26%T、18%G、23%C，解释这种现象。

9．大肠杆菌染色体长约 1 100μm。

A．大肠杆菌染色体有多少对碱基？

B．这条染色体双螺旋须旋转多少周？

10．两段 DNA 片段如下：

①TAGCTCC　　③GCTCCTA

②ATCGAGG　　④CGAGGAT

这两段 DNA 分子经热变性后再复性，如果（a）第 2 条单链与第 3 条单链复

性，(b) 第 1 条单链与条 4 条单链复性。图示复性后的 DNA 分子的结构，并标明序列和极性。

11. 直接用于蛋白质合成过程的遗传信息的形式是（选择正确答案）：

A. DNA　　B. mRNA　　C. rRNA　　D. 核糖体

12. Cot1/2 与基因组大小一致。

A. 对　　B. 错

13. Cot1/2 与基因组复杂性相一致。

A. 对　　B. 错

14. 滚环复制

A. 是细菌 DNA 复制的一种主要机制。

B. 扩增原来的复制子。

C. 始终使原来的复制子产生双链环状拷贝。

15. 根据 tRNA 的反密码子与密码子互作的摆动理论，阅读完 61 个有义密码，理论上至少需要多少种 tRNA?

16. 有一段多肽链为 Arg－Gly－Ser－Phe－Val－Asp－Arg，这段多肽链由下列 DNA 片段编码：

—GGCTAGCTGCTTCCTTGGGGA—

—CCGATCGACGAAGGAACCCCT—

(1) 哪一条链为转录的模板链?

(2) 标明这两条单链的极性。

17. 大肠杆菌经紫外光照射后会形成胸腺嘧啶二聚体，有哪两种修复机制可以修复这种损坏斑？比较这两种修复机制的异同。

18. 简述大肠杆菌中重组修复的过程。在何种情况下，细胞会通过重组修复来修复 DNA 中受损坏的区域?

第二章
基因的特征

经典遗传学指出，每一种生物都具有许多控制性状发育的遗传单位，这些遗传单位叫做基因。现在我们知道，基因是DNA分子中一段核苷酸序列，所以个体的基因组成总括起来就叫做基因型（genotype）。根据对红色面包霉（*Neurospora crassa*）氨基酸突变型的研究知道，这些营养缺陷型突变体缺少与某种氨基酸合成有关的酶，于是就产生了一个基因编码一种酶的结论。后来知道，许多酶都是由一条以上的多肽链组成；还知道，有些多肽如肽激素、核糖体蛋白等又都不具有酶的催化活性。因此，后来就把一个基因一种酶的理论修改成“一个基因编码一条多肽链”。但是这种提法也并不完全正确，因为有些基因并不指导蛋白质或多肽链合成，如编码rRNA和tRNA的基因以及其他一些不翻译成多肽链的RNA基因等。虽然一个基因一种酶的理论并不完全正确，但它对于我们深入认识基因的结构与功能作出了重大贡献。

第一节　基因的结构特征

一、基因的一般结构

（一）转录单位

基因的基本功能部分是编码某种特殊大分子化合物的核苷酸序列，这段序列叫做转录单位（transcriptional unit），RNA聚合酶可以将其转录成与之互补的RNA链，包括rRNA、tRNA和mRNA等。如果转录子是mRNA，通过翻译，就能编码出蛋白质或者组成蛋白质的多肽亚基。DNA分子中这种决定蛋白质分子中氨基酸序列的区域叫做可读框（open reading frame），它含有一系列编码氨基酸的三联体，不含任何终止密码。由于基因的最初产物是能够与其编码区互补的RNA序列，所以对基因的组成习惯上就以双链DNA分子中与模板链互补的一条链表示。

（二）基因间的间隔序列

在两个基因的编码区之间存在一些不编码的核苷酸序列，这类序列统称为基因间间隔序列（intergenic spacer），其长度变化很大，从一对碱基到数千对碱基不等。间隔序列中与某个基因相连的一部分序列通常叫做侧翼序列（flanking sequence）。由于间隔序列中某些部分对基因的功能具有某种特殊作用，所以有时基因5'端的这种序列叫做前导序列（leader）或5'非转录控制区，基因3'端的非编码序列就叫做拖尾（trail）或3'非转录控制区。但是，在许多生物的基因组中，有些基因之间并不存在间隔序列，有些则是两个或更多的基因首尾连接在一起，甚至有些基因的首尾部分还相互重叠。

在所有基因中，双链DNA分子只有一条链含有编码序列，但是许多生物的基因组中，有些基因位于DNA分子的一条链上，而另一些基因则位于另一条互补单链上。由于基因的阅读始终是按3'到5'方向进行，所以在两条互补链上的基因就具有不同的极性。

（三）基因的控制序列

基因在表达过程中，受到一系列具有特殊功能的、短的DNA序列的调节，其中有些控制基因的转录启始和终止过程，有些确定翻译过程中核糖体与mRNA结合，而另一些则与基因接受某些特殊信号有关，所以分子遗传学中就把这些控制和调节基因表达的特殊序列统称为控制序列（controlling sequence）（图2-1）。在这些因子中，有些位于编码序列上游并同基因的编码序列位于同一条DNA分子上，所以它们又叫做顺式作用元件（cisacting element）。

1. *启动子*　在基因的5'端即5'非转录控制区中有一段被RNA聚合酶所识别和同RNA聚合酶结合的序列，这段序列叫做启动子（promotor），它控制基因的转录启始过程。在大肠杆菌中，启动子含有两段结构保守序列，一段位于转录起点上游第10个碱基以上，其共有序列（consesus sequence）为TATAAT，叫做普里布诺框（pribnow box）。另一段位于这一段序列上游约25个碱基的位置，其共有序列为TTGACA，这段序列有时也叫RNA聚合酶识别位点。在真核生物基因中，启动子通常是指准确而有效地启始基因转录所需的任何DNA序列。对编码蛋白质的基因来说，启动子还包括位于转录起点（+1位）周围的帽子位点（cap site），-30位的TATA框，-40至-110位之间的其他上游因子（upstream element）和增强子（enhancer）等。

2. *增强子*　增强子首先见于病毒基因，后来在许多真核生物基因中都有发现。在某些病毒基因的启动子附近，有一种特殊的结构保守序列GGTGTGGAAATTTG，它可以显著地提高启动子启始基因转录的作用。增强子有以下

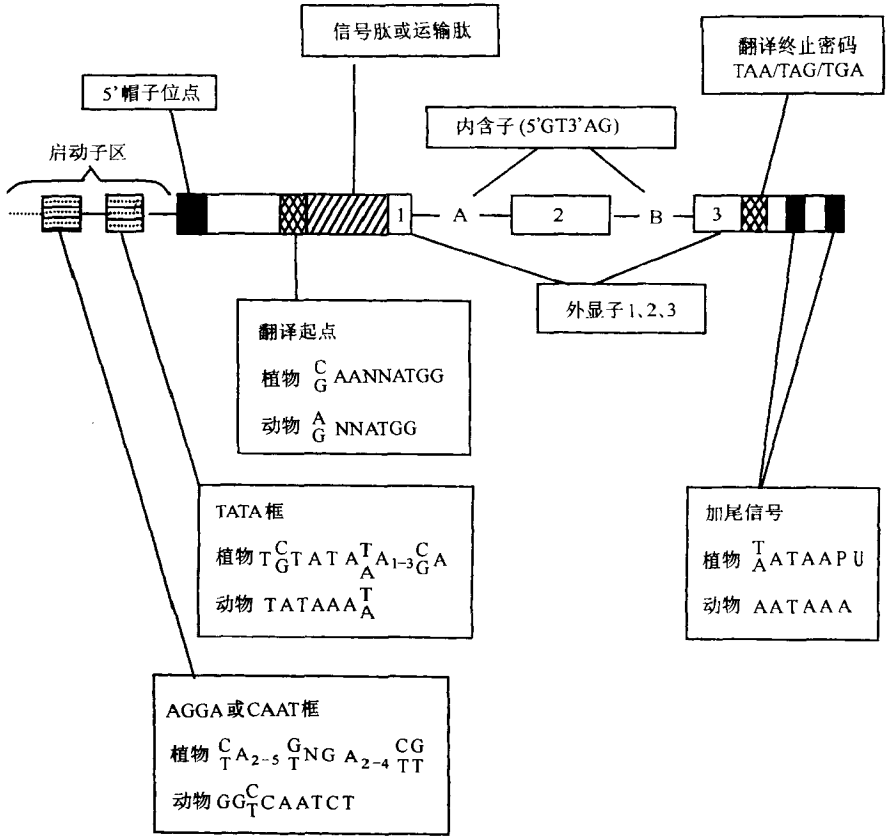


图 2－1　真核生物蛋白质基因的一般结构模型

几个特征：(1) 其增效作用具有双方向性，即它既可以作用于上游基因，也可以作用于下游基因，其作用范围可以远达与其相距 3kb 的基因。(2) 无论增强子与某个基因的极性相同或相反，对该基因都具有活性。(3) 增强子具有明确的组织和物种特异性。在猴肾病毒 SV40 中，增强子是 72bp 的重复序列的组成部分，但其中只有一个具有增效功能。

3. **核糖体结合位点**　在基因编码区的上游，除了存在上述非转录区中控制因子以外，在基因的转录区中，即在 mRNA 的翻译起始位点周围有一组特殊的序列，控制 mRNA 的翻译过程，其中主要为 ATG（mRNA 中为 AUG）起点密码及其前后的若干碱基。在细菌基因中，这段短的核苷酸序列叫做 SD 序列（Shine－Dalgarno sequence），同感序列为 AGGAGG，是核糖体的结合位点，它与 16srRNA3' 端 CCUCCU 结合。

ATG 密码的位置和可读框的方向并不是决定翻译启始过程的惟一因素。

例如将 AUG 插入到 mRNA 的前导序列中，这一密码对翻译的效应依赖于其侧翼序列。如果 AUG 前有 A（或 G）CC，其后有 G，这种序列组合对启始转录过程具有最高效率。在某些基因中，翻译起点密码为 GUG 和 UUG（见第四、五两章），虽然通常前者编码缬氨酸，后者编码亮氨酸，但作为起点密码，二者都编码甲硫氨酸。在大肠杆菌的氨甲酰磷酸合成酶基因中，翻译起点密码则是 AUU（通常编码异亮氨酸）。

4．*基因拖尾区中的信号序列*　基因 3’端非编码区中的主要信号是与终止转录过程有关的序列，叫做终止子（terminator）。根据对许多转录子的核苷酸序列分析，RNA 聚合酶终止转录的位点正好位于转录子本身的末端。转录终止子主要有两种类型，第一种终止子由一对富含碱基 GC 的颠倒重复序列以及寡聚 T 组成，这段序列可以形成发夹环结构（图 2－2）。可能，当 RNA 转录到达这一区域时，其自身通过碱基配对，形成发夹环而终止转录过程。第二类终止子目前还不很清楚，它们不富含 GC 碱基和寡聚 T 序列，但现在知道，细菌中这类终止子行使功能时需要 rho 因子和一种叫做 nusA 的蛋白质。值得指出的是，并非所有基因的终止子都相同，例如在酵母中，mRNA 有一段由 8 个碱基 TTTTTATA 组成的序列，当这段序列活化后，转录过程就终止在另一组序列 CAATATTTG 的最后一个 T 碱基上。

5．*加尾信号*　真核生物 mRNA 的 3’端都有一段多聚 A 尾巴（poly A tail），这种尾巴不由基因编码，而是在转录后加到 mRNA 上的。加尾过程受位于终止密码 3’端的加尾信号序列所控制。现在知道，动、植物基因都有一个以上加尾信号序列，动物基因的同感序列为 AATAAA，植物基因的加尾信号变化较大，为 ATAATAAApu。

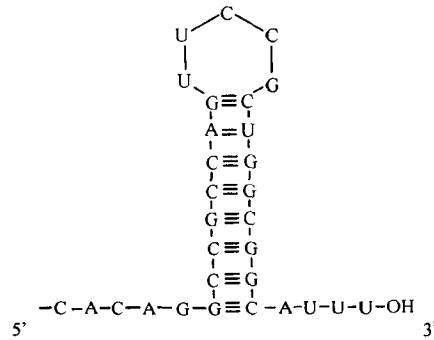


图 2－2　转录终止位点的结构

6．*内部控制信号*　在某些小 RNA 分子如 tRNA、5SrRNA 以及病毒的某些 RNA 基因中，启动子序列位于基因的编码序列之内。关于这种位置与启动转录之间的关系目前还不清楚。

（四）隔裂基因　在真核生物中，大多数基因的编码蛋白质的序列都由若干不编码的序列所隔开，编码区中这种不编码的部分叫做内含子（intron），编码的部分叫做外显子（exon）。基因在转录以后通过加工，将内含子去掉，然后再将外显子连接起来，形成一个连续的编码区（第四章），像这样一种基因叫做隔裂基因（split gene）。

隔裂基因中内含子的数目因物种和基因不同而有差别，有些基因只有一个内含子，有些则有多个内含子。例如小鼠肌肉 α－肌动蛋白基因（muscle α－actin gene）有 6 个内含子，酵母线粒体的细胞色素氧化酶亚基Ⅰ基因则有 7～9 个内含子。

隔裂基因在转录过程中，内含子和外显子一起转录，对某些基因来说，除掉内含子并不影响基因的正常功能和基因产物的加工；而对另一些基因来说，如 SV40 的基因、酵母 $tRNA^{Tyr}$ 基因和细胞色素氧化酶亚基 I 基因等，除掉内含子往往导致转录子不能进行加工和失去正常功能。

具有上述结构特征的基因都属简单基因，因为这些基因在基因组中都有相对固定的结构和位置，并且一个基因只编码一种产物。在各种生物的基因组中还存在一些结构和功能都不完整的基因，这类基因叫做假基因（pseudogene），有的缺少转录启始和终止信号，有的缺少完整的编码序列。下面就一些结构和功能都比较复杂的基因分别加以说明。

二、基因的变异类型

（一）二型蛋白基因

二型蛋白基因（diplomorphic gene）是指含有两个相互连接的编码区的基因，编码两种不同功能的多肽，其特征是 5’编码区的产物为信号肽（signal peptide）或运输肽（transit peptide）（第四、五章）。例如小鼠的激肽释放酶（kallikrein）基因的前体 mRNA 长 867 个碱基，其中前导序列为 24 个碱基，3’端拖尾区为 48 个碱基，编码序列为 795 个碱基，编码长 265 个氨基酸的蛋白质（图 2－3）。翻译的初级产物叫做前酶原（preproenzyme），其前面长 17 个氨基酸的亚基行使一种运输蛋白质的功能，叫做信号肽。在内质网膜上的蛋白质合成过成中，信号肽与一种叫做运输识别颗粒（transit recognition particle）的特殊受体互作，识别内质网的表面位点和将与之相连的激肽释放酶原运输过内质网膜。然后，在运输识别颗粒的帮助下，整个前酶原进入内质网泡腔中，最后通过酶切，将信号肽除掉，形成激肽释放酶原。激肽释放酶原分泌到细胞外表后，还需通过进一步加工，将前面的 11 个氨基酸除掉，才形成具有活性功能的激肽释放酶。因此，激肽释放酶基因编码两个功能部分，即信号肽和激肽释放酶，二者由一段不具活性的区域隔开。

（二）隐形蛋白基因

在真核细胞中，有些基因的具活性的最终编码产物隐蔽在前体蛋白质之中，当前体蛋白通过酶切或加工之后，才形成具有功能的蛋白质，编码这类蛋

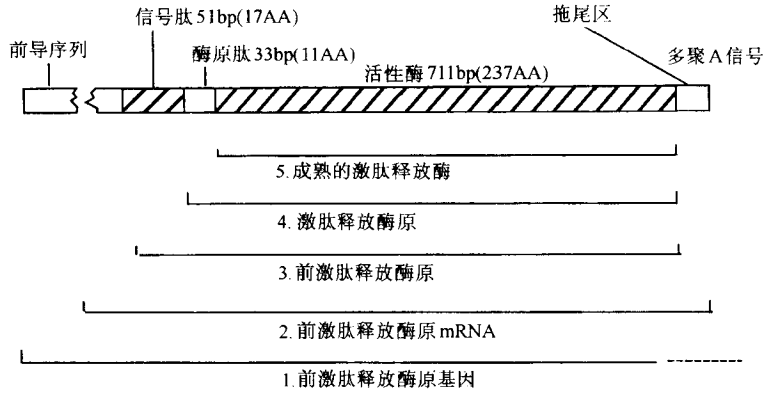


图 2－3　二型蛋白基因的结构

小鼠的激肽释放酶基因具有一段编码信号肽的前导序列，其后为一段编码抑制因子或肽原的序列，当这两段序列除掉以后，才能形成具有活性的激肽释放酶

白质的基因叫做隐形蛋白基因（cryptomorphic gene）。例如啤酒酵母（*Sacckaromyces cerevisae*）的交配外激素（mating pheromone）基因，也叫α－因子基因（MFα－1）就属于这类基因。交配外激素基因的转录子长 495 个核苷酸，翻译产物长 165 个氨基酸。这段肽链包括 4 段首尾连接的α－因子，每一段长 13 个氨基酸。这 4 个拷贝的α－因子之间均由一段短的前导肽隔开，除第一个前导肽长 6 个氨基酸外，其余均长 8 个氨基酸（图 2－4A）。前体蛋白质在翻译后以完整的形式贮存于细胞之中，当酵母细胞受到某种未知的内、外因素影响时，前体蛋白通过水解，释放出α－因子和前导序列。总的来说，隐形蛋白基因表达时，首先编码出一条大蛋白质分子，然后其靠羧基的半部分经过切割和加工，最后形成具有活性的交配外激素。前体交配外激素的氨基部分相当于二型蛋白基因的前半部分，其作用是使其实际的功能部分停留在不活动状态，当后半部分被利用时，氨基部分也随之被切割。至于其中的连接序列是否具有运输蛋白质的功能，目前尚不清楚。

高等真核细胞的隐形蛋白质基因比酵母交配外激素基因复杂得多，例如哺乳动物胰脏的胰高血糖素基因。胰高血糖素基因编码的初级翻译产物叫做前高血糖素原，其前面的 20 个氨基酸为信号肽。前体蛋白的主要部分为高血糖素原，长 160 个氨基酸（图 2－4B），包括高血糖素的结构序列和两条不同的叫做类高血糖素的多肽。前体蛋白在加工过程中，首先去掉类高血糖素多肽，形成长 69 个氨基酸的、叫肠高血糖素的激素，然后再通过第二次加工才形成具有功能的高血糖素。

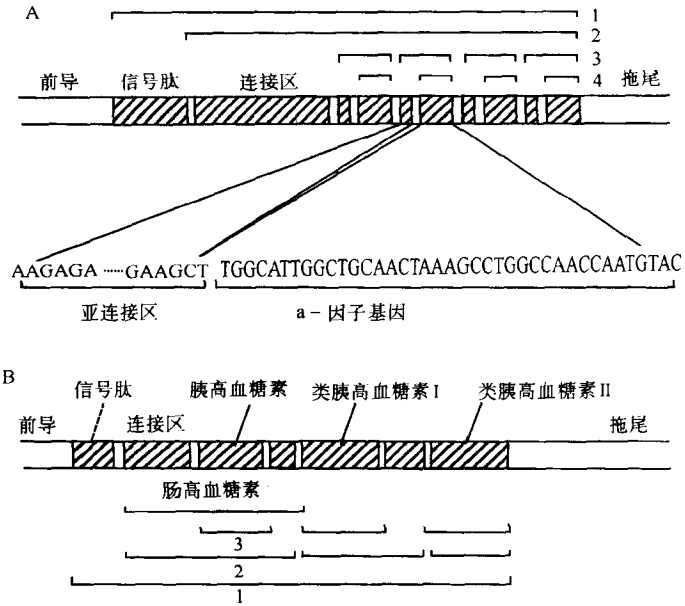


图 2－4　两种隐形蛋白质基因

A. 酵母性外激素基因　B. 哺乳动物胰高血糖素基因涂阴影区示初级转录子，但其中只有一部分编码最终的功能产物

（三）隐密基因

在原生动物的线粒体以及某些真核生物的线粒体基因组中，有些基因以一种节略的形式存在，它们缺少完整的蛋白质编码序列和 mRNA 分子中与翻译有关的其他一些功能结构。这类基因在转录后通过一种特殊的加工方式即 RNA 编辑（第五章），对 RNA 进行加工，将转录子变成可翻译的 mRNA，这类基因叫做隐密基因（cryptogene）。根据对锥虫（*Trypanosomatide brucei*）以及其他几种原生动物线粒体基因组的分析，细胞色素 b 基因（cyb）、细胞色素氧化酶亚基Ⅱ（coⅡ）和亚基Ⅲ（coⅢ）基因都是隐密基因。例如，在 *T. brucei* 中，cyb 基因在转录后，mRNA5' 端加进了 34 个 U，而在另外两种锥虫 *Leishmania tarentolae* 和 *Crithidia fasiculate* 的线粒体中，cyb 基因的 mRNA 则加进了 39 个 U。又如 *T. brucei* 的 coⅢ基因，其初级转录子在 158 个位点上共加进了 398 个 U，在 9 个位点上缺失了 19 个 U。隐密基因在转录后通过 RNA 编辑，或者是将一种不具功能的转录子加工成可翻译的 mRNA，或者是通过修改已具功能的 mRNA 分子而改变其编码产物的氨基酸序列。

（四）复合基因

复合基因（complex gene）主要为脊椎动物免疫系统的基因，其结构相当

复杂，现以免疫球蛋白（immunoglobulin）基因为例加以说明。

免疫球蛋白又称抗体（antibody），现在已知有5种类型的免疫球蛋白即A、D、E、G和M，其中有些还可分成几个亚种。一般说来，每种免疫球蛋白都由4个亚单位组成（图2-5），即两条重链（H链）和两条轻链（L链）；轻链又可分成κ（kappa）链和λ（lambda）链，所以免疫球蛋白的一般结构可用κ2H2或λ2H2表示。

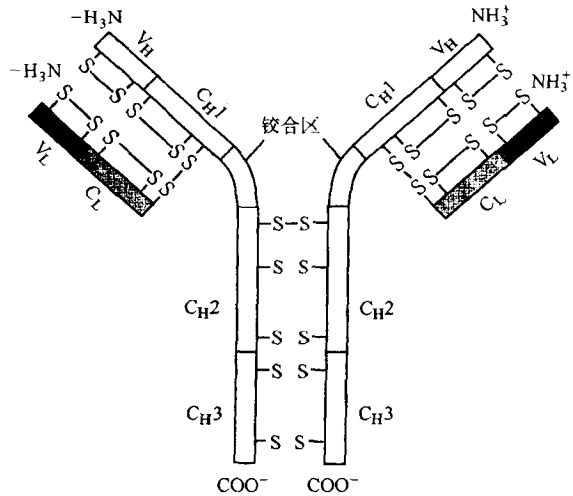


图2-5　免疫球蛋白的结构模型

两条轻链（阴影区）和两条重链（无阴影区）通过二硫键结合在一起，每一条重链和轻链的氨基末端部分都有一个可变域（V_H和V_L），轻链只有一个恒定域（C_L），而重链则有三个恒定域（C_H1、C_H2、C_H3）和一个铰链区

免疫球蛋白的两条轻链分别由212～214个氨基酸组成，每一条链又大至均等地分成一对相对应的区域。含有羧基末端的半部分通常变异较少，所以叫做恒定域（constant domain），氨基末端的半部分叫做可变域（variable domain），其变化范围很大，主要与抗体的特异性有关。

上述5种免疫球蛋白重链的变化都比轻链的变化更大。从结构上讲，重链与轻链相似，都由恒定域和可变域组成，但是重链有3～4个恒定域，而轻链κ和λ变异体都只有一个恒定域。在重链的恒定域中，前面的两个由一段短的铰链区（hinge）隔开。除了L链和H链以外，有些成熟的免疫球蛋白分子（IgA和IgM）还具有一条J链，其功能是将某些亚基连接在一起。在IgA中还有一个S链，这条链只在腺体中形成。

上述5种免疫球蛋白的主要差别在于恒定域不同，因此每一类型的恒定域的编码序列各不相同。恒定域的编码序列常用希腊字母表示，如IgM的恒定

域写为 Cμ，IgA 的恒定域写为 Cα，IgD 为 Cδ 等加以区别。又由于同一类型的免疫球蛋白有多个恒定域，所以各个恒定域的编码序列又可进一步用数字加以区别，如 IgM 恒定域的各个编码区分别为 Cμ1、Cμ2、Cμ3 和 Cμ4。此外，这 5 种免疫球蛋白的每一条链在基因组中都有多个拷贝的编码序列，所以其基因组成就非常复杂。

编码重链可变域的结构也是由几个不同部分组成。第一个区域为 V_H，编码重链可变部分的大部分肽链，其后为一条短的、叫做 D_H 的片段和几条叫做 J 的片段，J 片段最后能将可变域和恒定域连接在一起。免疫球蛋白基因的前导区也是一种复合结构，原始的前导区很长，其中只有一部分最后拼接在一起用于转录。

由于恒定域和可变域复合体的各个成员既不编码某个具体蛋白质的亚基，也不编码某种功能 RNA 分子，所以从逻辑上讲，免疫球蛋白基因的各部分并不符合基因的一般定义，但目前通常把组装成最终转录单位之前的可变域和恒定域的每一个成员叫做基因节段或基因子（genelet）。在最后形成的基因中，以重链可变域为例，V_H 和 D_H 基因子组合只与一个 J 基因子结合，究竟同哪一个 J 基因子结合，取决于细胞的类型、生理状态以及环境条件等一系列复杂因素。根据环境条件，重链可变域基因子在转录之前要进行一系列准备活动，所以 V_H 和 D_H 部分一方面有可能同 J3 结合，另一方面有可能在另一个细胞的另一个时期同 J1 和 J 组的其他成员结合。通过这样一种组合方式，就形成编码 V_H 域的前基因（pregene）（图 2－6）。

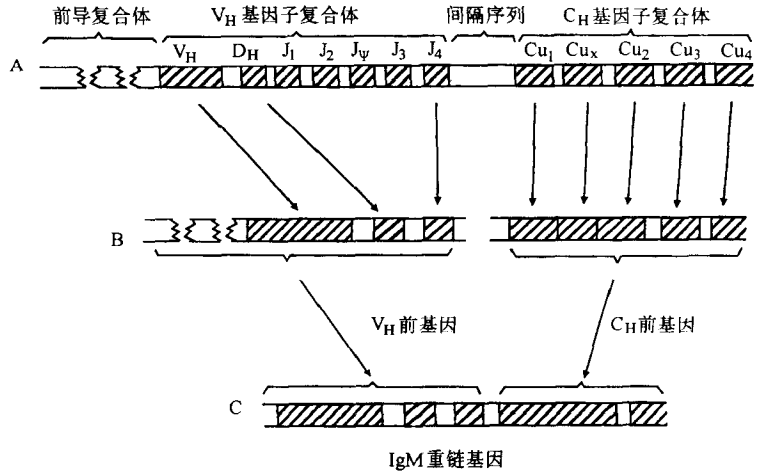


图 2－6　免疫球蛋白重链基因结构及其组合过程

A. 基因子复合体　B. 降解成前基因　C. 组装成 IgM 重链基因

同样，重链恒定域复合体也可能通过相似的机制降低各个基因子之间的间隔片段的长度而形成前基因。有关轻链和重链功能基因的组装详见第七章。

在重链编码区中，只有当这两种前基因结合在一起以后，其前导序列、肽链编码区和间隔序列才能用于转录。免疫球蛋白基因在制备转录单位过程中这一系列重要的、具有高度选择性的步骤可能是在各种内、外因素影响下，基因组中其他各种有关基因共同作用的结果。在上述 5 种免疫球蛋白中，细胞究竟合成哪一种，完全由细胞内的活动所决定。例如 IgM 首先出现在许多幼期细胞中，在成熟细胞中 IgM 停止合成，而合成第二类免疫球蛋白（通常为 IgG）。

第二节　基因的组织

一、操 纵 元

操纵元（operon）一词首先来自对细菌突变体的分析。在细菌中，编码若干种功能上有关的酶类的基因在转录时受一个开关单位控制，所有基因作为一个单位转录，形成一条多顺反子 mRNA，各个顺反子之间由一段短的间隔序列隔开（图 2－7）。像这样一种“超级基因”就叫做操纵元或操纵子。由于转录子在转录后迅速翻译，所以就有可能等量合成几种在功能上有关的酶，这种蛋白质合成过程就叫做协同控制（coordinate control）。当然，并非所有多顺反子 mRNA 的翻译都是协同控制，有些操纵元的上游顺反子的翻译效率高于下游顺反子，而另一些操纵元的下游顺反子的翻译效率则可能高于上游顺反子。

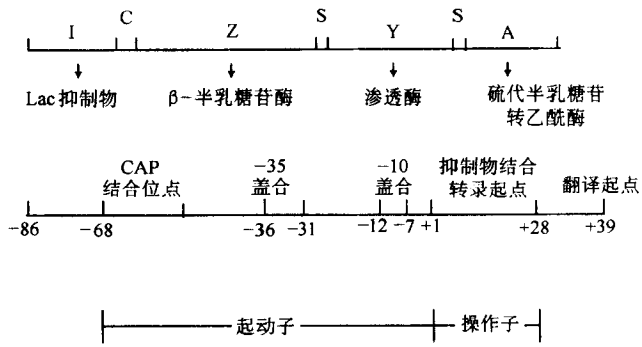


图 2－7　乳糖操纵元的结构

上图为整个操纵元的结构，其中 I、Z、Y、A 分别为结构基因符号，操纵元之下为这些基因相应的编码产物；C 为控制区，S 为间隔序列。下图为控制区的详细结构，数字示某个核苷酸距转录起点核苷酸的距离，转录起点上游的核苷酸以负数表示

有些操纵元的表达是相当稳定的，也就是说由其结构基因编码的若干种酶能够稳定地合成，较少受到细胞内、外其他因素的影响，这类操纵元叫组成型操纵元（constitutive operon）。还有一些操纵元的表达则受到严格控制，它们受细菌细胞中某些代谢产物的调节。如果某种代谢产物能够使操纵元表达，这种操纵元叫诱导型操纵元（ inducible operon）。如果操纵元的表达受到细胞中某些代谢产物的抑制，这种操纵元叫抑制型操纵元（repressible operon）。细菌中还有一些在功能上相关，参与同一代谢活动的多个基因分布在几个不同操纵元中，但各个操纵元的表达都受同一种活化蛋白的调节。像这样一组受同一种活化物调节的所有操纵元就叫做一个调节元（regulon）。

一个操纵元的表达涉及两类基因，即结构基因（structural gene）和调节基因（regulatory gene）。结构基因编码某些稳定的 RNA 或 mRNA；而调节基因则编码与调节结构基因转录有关的产物，所以从这个意义上讲，调节基因也属于结构基因的范畴。在操纵元中，在结构基因上游还存在启动子和另一叫做操作子（operator）的区域。同启动子一样，操作子并不编码任何产物，只是对其下游结构基因起着调控作用，所以操作子不是一个基因，不具有功能互补性。在有些操纵元中，操作子与启动子紧邻，而在另一些操纵元中，操作子则与启动子重叠（第三章）。调节基因并不一定与其所控制的操纵元位于同一染色体上，它有可能位于不同复制子如质粒上，所以就把调节基因或由其编码的产物叫做反作用子（trans－acting element）。由于启动子和操作子位于它们所控制的基因的上游，同在一条染色体上，所以二者都称为顺式作用元件（cis －acting elemnent）。反作用子发生突变后，可以通过引入载有未发生突变基因的质粒来校正，即在功能上可以由未发生突变的基因所互补。而顺作用元件发生突变后，则不能通过这种方式加以纠正。

并非所有操纵元都编码蛋白质，例如，在细菌和真核生物基因组中编码三种 rRNA（细菌为 23s、16s 和 5s）的基因在转录时也作为一个单位进行转录，所以也叫做操纵元。在大肠杆菌中，因菌株不同，现在已知有 7 种不同组织的 rRNA 操纵元。所有 rRNA 操纵元的总体组织都基本相似（图 2－8）。每一类操纵元之前都有一段前导序列，其中载有识别信号或类似结构。前导序列下游为三个编码区，分别编码 16s、23s 和 5s rRNA。在前两个编码区之间还有一段间隔序列，其中含有 1～2 个 tRNA 基因。每一操纵元都转录出一条前体 RNA，然后通过切割，产生各种成熟的 rRNA 和释放出 tRNA。所有 rRNA 操纵元都有 2 个启动子，第一个启动子 P1 位于 16SrRNA 起点上游约 300bp 的位置，在 P1 下游约 110bp 的位置有第二个启动子 P2，可能 P1 是主要的启动子。真核生物如果蝇的 rRNA 操纵元编码三种 rRNA，其顺序为 18S－5.8S－26S。

同原核生物相似，各种 rRNA 编码区之间都有一段被转录的间隔区，但不同的是间隔区不编码 tRNA。在哺乳动物中，rRNA 操纵元（18S－5.8S－28S）首先转录出一条 45S 的前体 RNA，然后通过几个连续切割步骤，释放出 18S、5.8S 和 28S 的 rRNA。

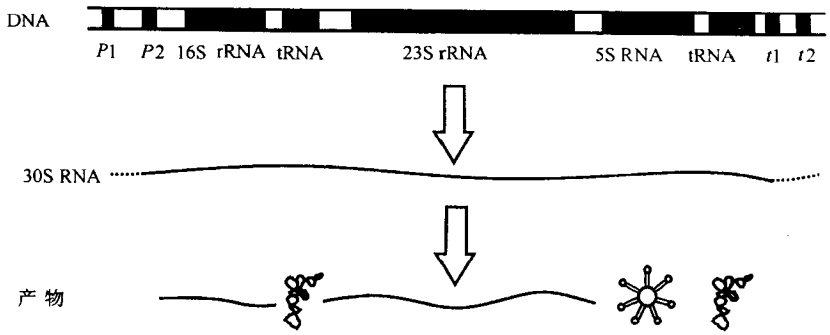


图 2－8　原核生物 rRNA 操纵元的组织及其编码的产物

操纵元的每一编码序列都由被转录的非编码序列隔开；前导区和拖尾区的长度因所使用的启动子和终止子不同而有差别。前体转录子通过酶切后形成各种 rRNA 和 tRNA

二、重复基因与基因家族

（一）首尾连接的重复基因

在各种生物的基因组中，许多基因都是单拷贝，没有重复序列。如各种病毒基因、大肠杆菌的 $tRNA^{Trp}$ 基因以及各种细菌的赖氨酸操纵元等都是单拷贝基因。在真核生物的基因组中，单拷贝基因更为普遍，如果蝇的 α_4^- 微管蛋白（tubulin）基因、鸡的 α_2I 型胶原蛋白（collagen）基因以及红色面包霉（*Neurospora crassa*）的 H_3 和 H_4 组蛋白基因等也都是单拷贝基因。

但是在各种生物包括细菌、酵母以及高等动、植物的单倍体基因组中，许多基因都有两个或两个以上拷贝，它们彼此相同或十分相近，这样一组基因就叫做基因家族（gene family）。一般认为，基因家族里的各个成员都是由其祖先的核苷酸序列发生重复而形成的。在生物的进化过程中，一些重复基因仍旧保持在一起，成为一个基因簇（gene cluster），例如爪蟾（*Xenopus*）的 18S、5.8S 和 28S rRNA 基因各有 450 个拷贝，它们都集中在核仁组织者周围，形成一个 rRNA 基因簇。在有些基因簇中，各个重复基因首尾衔接在一起，如人体和果蝇的 5S rRNA 基因就属于这种范畴。真核细胞中的 5S rRNA 基因与前

述三种 rRNA 基因的操纵元并不衔接在一起，而是位于基因组的其他位置。在人基因组中，多拷贝的 5S rRNA 基因衔接在一起，而在果蝇中，约 160 个 5S rRNA 基因分两组衔接排列。又如人的绒毛膜促性腺激素（chorionic gonadotropin）的 β 亚基基因也为衔接排列的重复基因，至少有 8 个成员衔接在一起，其中有 4 个成员为顺式衔接重复（tandem repeat），其余为颠倒重复（inverted repeat）。颠倒重复基因可能是这类重复基因的一种变异形式。

（二）分散排列的重复基因

有一些重复基因并不衔接排列在一起，而是分散在整个基因组中。例如，果蝇的微管蛋白基因有 4 个成员，虽然其中有三个基因都与 α4 基因不同，但这三者之间都有一定序列同源性，它们分散于果蝇基因组的不同位点上。又如酵母的蔗糖酶（invertase）基因也有 7 个成员，也分散在基因组中，其中有些成员还编码两种以上的酶。在果蝇中，研究最为清楚的为 copia 基因，（见本章第四节），在果蝇不同组织的细胞中，其转录子都极为丰富，约占多聚 A 化 mRNA 总量的 3%，这一基因家族的各个成员也都分散在基因组的不同位点上。在酵母中，另一类研究较详细的基因为 4 种组蛋白基因（图 2－9），酵母不含 H_1 组蛋白，在这种四种基因中，H_2A 和 H_2B 衔接在一起，H_3 和 H_4 衔接在一起，但这两类衔接的基因具有不同的转录极性，并且这四个基因在基因组中都只重复一次。据分析，两个 H_2B 基因的产物只在 4 个位点上的氨基酸有所不同，而其 mRNA 则在 390 个位点上有 400 个核苷酸不同。这种现象表明，重复基因并不一定相同。前述大肠杆菌 rRNA 操纵元的各个重复单位也都不完全相同的。

（三）基因簇的组织

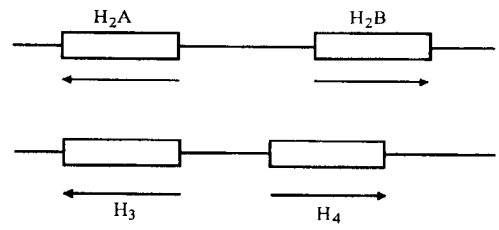


图 2－9　酵母的 4 种组蛋白基因，箭头示转录方向

在有些基因簇中，尽管其中各个基因的结构差别很大，但它们的编码产物都参与同一代谢过程或具有相似功能，前述组蛋白基因就属于这一类型。组蛋白基因是研究最早的基因之一，因为各种组蛋白在细胞中含量特别丰富，它们在细胞分裂之前的某个特殊阶段合成。除了在酵母和某些脊椎动物中组蛋白基因只有少数几个拷贝外，在其他许多生物中组蛋白基因都是多拷贝基因（表 2－1）。在无脊椎动物特别是在海胆和果蝇中，5 种组蛋白基因各以一个拷贝衔接在一起，组成一个基因簇，然后各个基因簇又以重复单位的形式衔接在一起。每一重复单位中的各个基因通常都没有

内含子，都独立地转录，其 mRNA 也不多聚 A 化。

表 2－1　组蛋白基因簇的重复数及其组织

物　种	重复频率	重复长度（kb）	组　　织
海胆	300～600	5～7	$H_1-H_4-H_2B-H_3-H_2A$
果蝇	100	25	$H_3-H_4-H_2A-H_2B-H_1$
爪蟾	20～50	5.8	$H_3-H_4-H_2A-H_2B$
			$H_3-H_4-H_2A-H_2B-H_1-H_3-H_4$
鸡	10	710	多样
人体	10～20	710	多样

在不同生物中，组蛋白基因簇中各个成员的排列顺序各不相同。在果蝇中，5 种组蛋白基因的排列顺序通常为 $H_1-H_3-H_4-H_2A-H_2B$，但在某些情况下排列顺序为 $H_4-H_3-H_1-H_2B-H_2A$。在爪蟾（*Xenopus laevis*）中，组蛋白基因簇的各个成员的排列顺序通常为 $H_1-H_2B-H_2A-H_4-H_3$，但也有少数变异体按果蝇中第二种顺序排列。在另一种爪蟾（*X. tropicalis*）中，各种组蛋白基因的顺序依次为 $H_4-H_1-H_3-H_2A-H_2B-H_4-H_1$，其中有两个 H_1 基因变异体。在海胆中，各个组蛋白基因则依次为 $H_1-H_4-H_2B-H_3-H_2A$。在某些真核生物如海胆和滑爪蟾中，各个组蛋白基因的转录方向都相同，而有些生物如果蝇的 H_4 基因则与其他 4 个基因的转录方向相反。在所有生物中，各个组蛋白基因之间都有一段间隔序列。

根据对海胆的研究，组蛋白基因是一类与发育阶段有关的基因，其拷贝数与不同发育阶段表达有一定相关性。例如在囊胚阶段表达的 α- 亚型组蛋白基因在基因组中有数百个几乎相同的衔接重复，每一拷贝都编码同一种组蛋白，这种现象叫做基因扩增（gene amplification）。而在原胚阶段或在胚胎发育后期表达的基因重复次数则低得多，每一基因簇或由 H_1、H_2A 和 H_2B 组成，或由 H_3 和 H_4 基因组成，但这两种基因簇都分散于基因组中，每两个基因簇之间通常相距数千个碱基。又如，普通果蝇的卵壳蛋白基因编码 20 种以上不同蛋白质，在卵细胞中这些基因均为单拷贝，而在卵泡细胞（follecle cell）中，在卵子发生阶段扩增成数千个拷贝。在这些卵壳蛋白基因簇中，其中有一个基因簇位于 x 染色体上，在卵子发生初期，该基因簇扩增 14 至 16 倍，呈衔接排列。另一个基因簇位于第三染色体上，包括 4 个卵壳蛋白基因，它们在卵壳形成后期活动，并扩增约 60 倍，形成衔接重复基因簇。

三、重叠基因

重叠基因（overlapping gene）是指两个或两个以上基因共用一段 DNA 的现象，最早见于病毒基因组。例如，在一种单链 DNA 噬菌体 φX174 中，根据这种噬菌体的基因产物推算，其基因组应有 6 000 个以上核苷酸，但根据 DNA 序列分析，实际上不到 5 400 个核苷酸，其原因就是许多基因相互重叠。后来发现，重叠基因普遍存在于线粒体（mitochontron）、叶绿体（chloroplast）、真核生物的细胞核和原核生物基因组中。

重叠基因可以分成三类 7 种（图 2－10），其中第一类重叠基因包括几个极性相反的基因，也就是说它们分别位于双链 DNA 的两条单链上。第二类重叠基因是指位于同一条 DNA 单链上两个基因，因此各个基因极性相同。第三类重叠基因是指重叠区域含有三个或更多的基因。

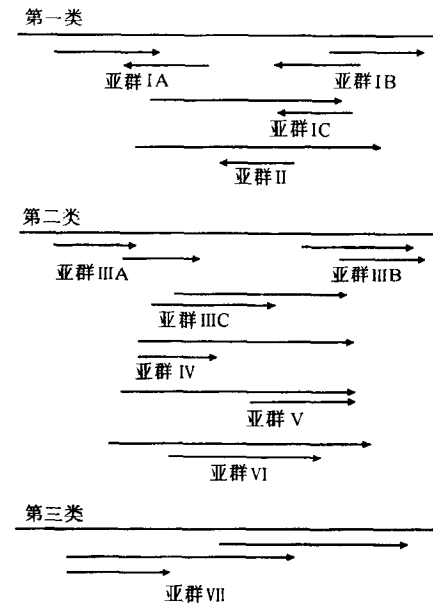


图 2－10　三类 7 种主要的重叠基因

第一类重叠基因有两个亚群，其中亚群 I 基因只是在末端部分重叠。根据重叠部分的长短和部位不同，又可分成 5’端重叠、3’端重叠和一个基因 5’端与另一基因 3’端重叠三种情况。亚群 II 是指一个较小的基因位于另一较大基因的编码区内，但二者分别位于相对的 DNA 单链上，所以含有这类重叠基因的生物体必须是双链基因组，它既可以是双链 RNA，也可以是双链 DNA。

第二类重叠基因可分成 4 个亚群，即亚群 III、IV、V、VI。亚群 III 基因又有三种变异型（IIIA、IIIB 和 IIIC），在 IIIA 中，一个基因的 5’端正好位于另一基因的 3’端。在 IIIB 中，一个较短基因的 3’端略超出另一较长基因的 3’端；而亚群 IIIC 则是一个较短基因的 5’端略超出另一较长基因的 5’端。在亚群 IV 基因中，两个重叠基因共用同一起点密码，但其中有一个基因较短。亚群 V 重叠基因从两个不同位点启始转录，但在同一位点终止转录过程。亚群 VI 则为一个较小的基因完全位于另一较大基因之内。

第三类重叠基因即亚群 VII 基因，这些基因的重叠区涉及三个或更多的基

因。

根据对许多种病毒基因组的研究，亚群Ⅲ基因是病毒中最普遍的重叠基因，其中一个基因的转录起点直接与另一基因的转录终止区重叠。例如在 T_4 噬菌体基因组的下游有一个编码胸苷酸合成酶（thymidylate synthase）的基因（ts）和一个二氢叶酸还原酶（dihydrofolate reductase）基因（frd），这两个基因同属一个转录单位，二者的编码区只有一个密码重叠。

T_4 frd － － － GAA TCA GTA TAT AAA TGA

T_4 ts － － －　　A TG A AA'C AA'T AC'C AA'G AT － － －。

在翻译过程中，当 frd 顺反子翻译时，位于其 3' 端的 ts 顺反子的编码区则受到破坏。相反，如果 ts 基因翻译，frd 顺反子的编码区则受破坏。

第三节　基因突变与重组

一、基因突变

一般说来，后代与其双亲在某些特征方面是非常相似的，这种现象说明 DNA 分子的组成是相当稳定的。但是，根据对各种生物的研究，DNA 分子的碱基序列确实能够发生变异，尽管这种变异的频率很低，但它是物种进化的基础。这些变异统称为突变（mutation），其发生可归纳为几种不同机制。

基因突变是指编码某种蛋白质或 RNA 的 DNA 序列中单个碱基发生改变的现象，包括碱基转换（transition）、颠换（transversion）、插入（insertion）和缺失（deletion）。转换和颠换造成点突变（point mutation），单个碱基对的插入和缺失则可使编码区发生本质改变。

（一）转换和颠换

转换是同类碱基之间的替换，即一种嘧啶碱基被另一种嘧啶碱基取代，或一种嘌呤碱基被另一种嘌呤碱基替换的现象。DNA 分子中碱基的互变异构化（tautomerization）、碱基类似物的替换以及亚硝酸的作用都可以引起碱基转换。如果碱基转换发生在某个基因的编码区中，其结果只会使单个密码发生改变，从而使蛋白质分子中某个氨基酸基团改变。

DNA 和 RNA 分子的嘌呤碱和嘧啶碱都可以发生互变异构化，例如常见的腺嘌呤碱为氨基形式，这种腺嘌呤通常同胸腺嘧啶配对，但其异构体亚氨基腺嘌呤则可以与胞嘧啶配对（图 2－11）。因此 DNA 分子在复制过程中，由 A－T 碱基对转变成 G－C 碱基对。

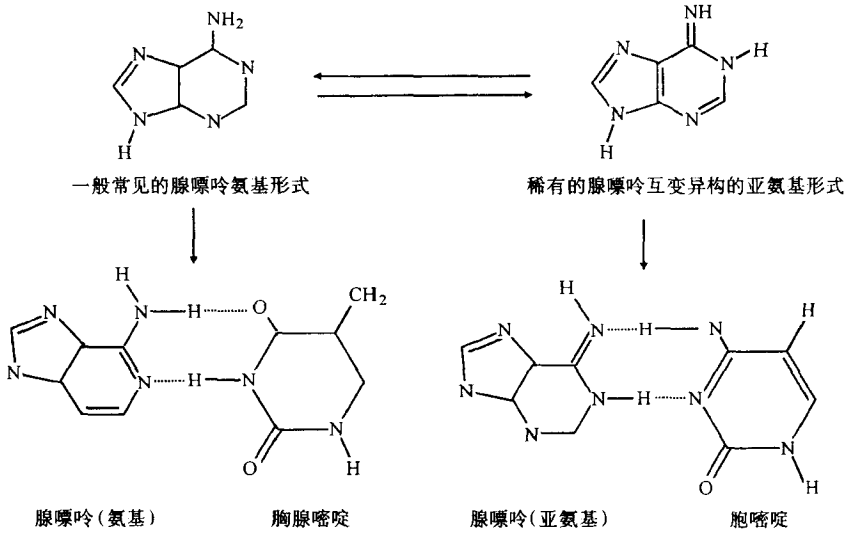


图 2－11　腺嘌呤的互变异构体及配对的碱基

颠换是指不同类型碱基之间的替换，即一种嘌呤碱替换一种嘧啶碱或者相反。颠换可以自然发生，也可以由化学诱变剂引起，但自然发生的颠换目前尚无直接证据。根据对正常人体血红蛋白（hemoglobin）和镰形细胞血红蛋白的分析，后者的形成可能就是碱基颠换的结果。这两种血红蛋白只有一个氨基酸的差别，正常血红蛋白在某个位点上有一个谷氨酸，而镰形细胞血红蛋白则在同一位点有一个缬氨酸。确定谷氨酸的密码为 GAA 或 GAG，而缬氨酸密码则为 GUU、GUC、GUA 和 GUG，镰形细胞血红蛋白中缬氨酸的出现可能是其密码中一个碱基 U 替换了 GAA 或 GAG 中间一个碱基 A 的结果。这是解释由正常血红蛋白到镰形细胞血红蛋白的一种最简单的假说。

DNA 分子中个别碱基的改变对多肽链中氨基酸序列具有不同影响。如果某个密码的第一个碱基或第二个碱基发生转换或颠换，就会使原来编码某种氨基酸的密码改变为编码另一种氨基酸的密码，或不编码任何氨基酸的密码，前者称为错义突变（missensemutation），后者称为无义突变（nonsense mutation）。例如，编码谷氨酸的密码 CAG 的第一个碱基颠换成 A 时，就会使原来的密码变成编码赖氨酸的 AAG；如果碱基 C 转换成 T，就形成终止密码 TAG，造成无义突变。如果发生改变的碱基位于某个密码的第三位，大多数情况下造成所谓的同义突变，因此多肽链中氨基酸序列就不会发生变化，许多简并密码就属这种类型。

在某些情况下，基因中的碱基对发生变化后，虽然使 mRNA 的密码子发

生变化，从而使蛋白质中氨基酸发生替换，但是该蛋白质的功能却不改变。这类突变叫做中性突变（neutral mutation）。例如，密码子 AGG 变成 AAG 会使赖氨酸变成精氨酸，由于这两种氨基酸都是碱性氨基酸，而且在特性方面都很相似，所以该蛋白质的功能就不会发生明显变化。

（二）插入和缺失

插入是指 DNA 链中插入一对或少数几对碱基，使基因的功能或可读框发生变化的现象。如果插入位点正好在某个基因的蛋白质编码区的 5’端，那么插入位点之后的一系列密码将发生错误，使原来编码某种蛋白质的编码区编码另一种完全不同的蛋白质，这种现象叫做移码（frame shift）。如果插入三个额外碱基，插入位点之后的密码序列则不会改变（图 2－12）。如果插入片段很长，则可能造成基因失活（见本章第三节）。

正常基因　　CAT CAT CAT － － － － － － 功能蛋白
　　his　his　his
插入一个碱基 CAT TCA TCA TCA T　　非功能蛋白
　　his　ser　ser　ser
插入三个碱基 CAT　[TAT]　CAT CAT － － － 功能蛋白
　　his　tyr　his　his

图 2－12　碱基变化与三联密码

插入一个碱基造成移码，插入三个碱基则插入位点以后密码序列不变

缺失是指 DNA 链中丢失一对或几对碱基。同插入一样，基因的编码区中缺失个别碱基也会造成移码（图 2－13）。

正常　Hb　　Thr Ser　　Lys　Tyr　Arg　Ter
　　ACN UC [U] AAA UAC CGU UAA GCU GGA GCC UCG GUA G

突变　Hb　　ACN UCA AAU ACC GUU AAG CUG GAG CCU CGG UAG
　　Thr Ser　Asn　Thr　Val　Lys　Leu　Glu　Pro　Arg　Ter

图 2－13　血红蛋白基因编码区中单个碱基缺失，造成合成一条较长的 α 链

图中示正常和突变 α 链 mRNA 的密码及其编码的氨基酸。方框内为缺失碱基

细胞中有一类移码抑制因子即一种特殊的 tRNA，这种 tRNA 可以识别由 4 个以上碱基组成的密码。如细菌移码抑制 $tRNA^{Lys}$ 能够识别 AAAA 或 AAAU，而不能识别正常密码 AAA。还有一种抑制子 tRNA 能够识别任何开头为 AAC 的由 4 个碱基组成的密码。

在某些噬菌体和病毒中，移码是一种正常现象，它调节蛋白质合成过程是

继续进行或是终止合成。在噬菌体 MS2 中，有一种移码可使核糖体识别一个裂解基因的可读框上游的一处终止密码，然后这种核糖体再识别裂解基因的起点密码。如果这一核糖体不终止翻译，它就阅读裂解基因的起点密码。所以这种依赖于移码的终止事件是启始裂解基因的蛋白质合成的前提。又如在酵母的 Ty 转座子的两个可读框中，当第一个顺反子 tya 的翻译在终止密码上终止后，核糖体可以通过识别移码，阅读第二个可读框 tyb。

上述现象表明，在终止密码上的移码往往导致翻译过程的通读（readthrough），终止密码 3' 端的一个碱基影响终止过程和移码的相对频率，因此从而影响终止信号的效率。

缺失和插入可以自然发生。一般认为，DNA 在复制过程中的错误可能是造成一、两对碱基发生缺失和插入的原因。例如下列 DNA 分子在复制过程中可能会随机地整合一个碱基 T，那么在下一轮复制过程中，这段序列就会多出一个 A－T 碱基对。

－A－A－A－A
－T－T T－T
　　＼／
　　T
－T－T－T－T
－A－A－A－A

较大片段的缺失和插入可能由另一些不同机制造成。在细菌中，有些缺失序列的末端都接有一段短的重复序列，缺失发生后，只是丢失了其中一些重复序列或间隔序列。插入因子（insertional element）或转座子（transposon）也会造成插入和缺失突变（见下节）。转座子通过一种目前还不十分清楚的机制将其自身或一个拷贝的序列转移到染色体的其他位点上，造成插入突变，同时也在其两侧造成几个碱基的缺失突变。

（三）DNA 复制机器引起的突变

在第一章中已经指出，DNA 复制是非常准确的，但复制过程也会不可避免地出现差错，从而导致某些突变。

大肠杆菌中有一种突变菌株叫做增变菌株（mutator），其 DNA 复制过程中出现的差错比正常菌株高出许多倍，所以造成基因突变的频率就比正常菌株高出许多。现已将这种增变菌株的突变定位到几个不同 mut 座位或称增变基因（mutator gene）上，其中几种主要基因的编码产物直接或间接参与除掉特殊损坏斑的过程。mut H、mut L、mut U 和 mut S 影响复制后修复系统的几种

成分，这种效应同 dam 甲基化酶基因发生突变后对误配对修复系统的影响相似。dam 突变体不能使 GATC 序列中腺嘌呤甲基化，所以误配对修复系统就不能区别模板链和新合成链，从而导致较高频率的自发突变。

mut D 基因突变后则使 DNA 复制过程中出现的差错得不到校正。mut D 编码 DNA 聚合酶Ⅲ全酶的ε亚基，这一多肽使聚合酶Ⅲ具有 3'→5' 外切核酸酶活性。如果聚合酶Ⅲ缺少这种外切核酸酶活性，新合成的 DNA 就含有大量突变。

mut Y 基因的编码产物与 A－G、A－C 误配对修复有关，它从误配对碱基对中除掉 A，所以这一基因发生突变就会导致 G－C→T－A 的颠换。mut M 增变基因也导致 G－C→T－A 颠换，mutT 则提高 A－T→C－G 颠换的速率。

在高等植物玉米中有一种增变基因叫做 Dt 因子，这一基因特异性地作用于叫做 a 的一个不连锁的座位。a 基因与紫色花色素苷合成有关，Dt 特异性地作用于 a_1，提高其突变成 A1 的频率，从而合成紫色花色素苷。$DtDta_1a_1$ 纯合体在无色籽粒的表面形成紫色斑点，可能每一斑点都代表一个能合成花色素苷的细胞克隆。突变体植株在茎、叶上也有紫色条纹。但目前对 Dt 因子只特异性地作用于 a_1 基因的机制还不清楚。

（四）化学诱变

某些物理因素和化学因素都可以诱发基因突变。物理因素只限于各种电离辐射和非电离辐射，由于其诱发基因突变的随机性很大，且会导致各种染色体结构变异，这里不作详细讨论。用于诱发基因突变的化学试剂较多，可以分成三大类，即碱基类似物、碱基修饰剂和嵌入剂。

1. *碱基类似物*　碱基类似物（base analog）是一类在分子结构上非常类似 DNA 分子中正常碱基的化合物。它们在 DNA 复制过程中常常与一个以上的正常碱基配对，从而导致突变。现以两种常见的碱基类似物 5－溴尿嘧啶（5BU）和 2－氨基嘌呤（2AP）为例加以讨论。

突变剂 5－溴尿嘧啶以两种状态存在。正常状态的 5BU 类似胸腺嘧啶，只与 DNA 中的腺嘌呤配对。在稀有状态下，它只与鸟嘌呤配对。当 5BU 被吸收到 DNA 中后，通过两种形式的互变而诱发突变（图 2－14）。这种情况与 DNA 复制错误而产生的自发突变相似。差别是碱基类似物通常以稀有状态存在，因此，由碱基类似物诱发突变的频率比自发突变的频率高得多。如果 DNA 分子中吸收了正常状态的 5BU，5BU 则与腺嘌呤配对。如果在 DNA 复制期间它转变成稀有状态，则可与鸟嘌呤配对。在下一轮复制过程中，G－5BU 碱基对将变成 G－C 碱基对，取代原来的 A－T 碱基对，造成转换突变，即从 AT 碱基对变成 GC 碱基对。如果在 DNA 复制过程中最初吸收的是稀有

状态的5BU，当它转变成正常状态后，就可诱发GC到AT碱基对的转换。因此由5BU诱发的突变可以通过5BU第二次处理突变体，使其发生回复突变。

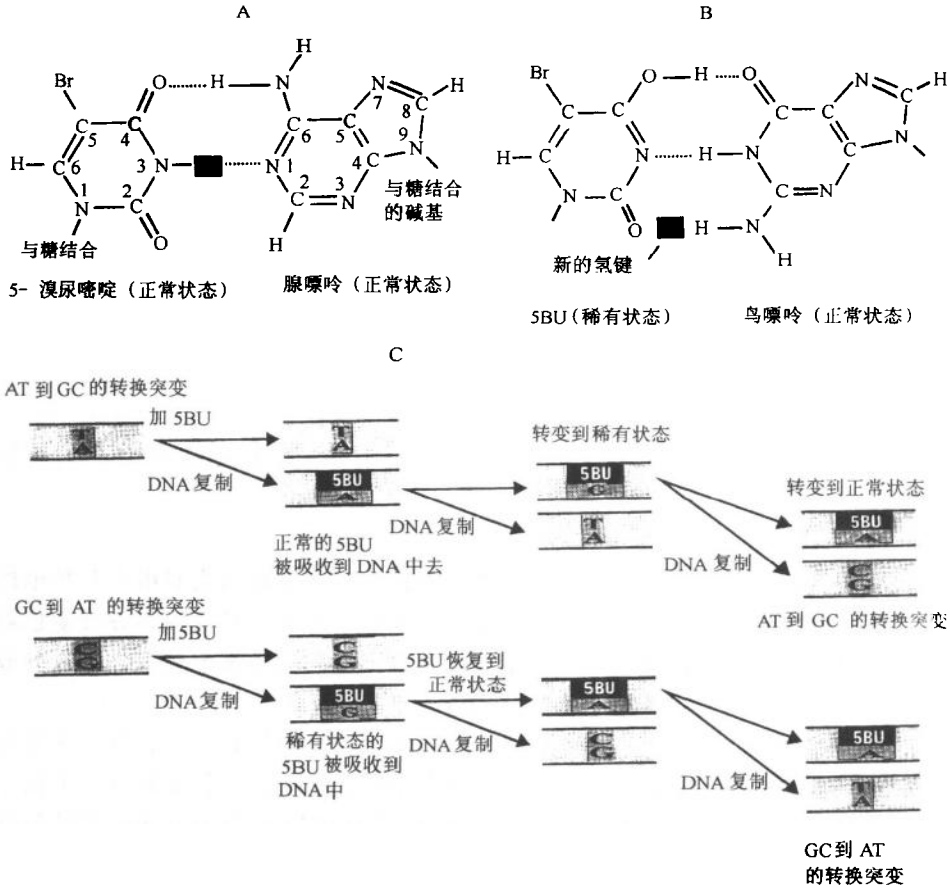


图2－14　碱基类似物5－溴尿嘧啶（5BU）的突变效应

A. 正常状态的5BU与腺嘌呤配对　B. 稀有状态的5BU与鸟嘌呤配对　C. 5BU诱发突变的两种可能机制

碱基类似物2－氨基嘌呤诱发突变的方式基本上与5BU相同（图2－15）。2AP也以正常状态和稀有状态存在。正常状态的2AP类似腺嘌呤，但与腺嘌呤不同的是在嘌呤环上不同位置有一个氨基。正常状态下，它与胸腺嘧啶配对。稀有状态的2AP类似鸟嘌呤，它与胞嘧啶配对。同5BU一样，它诱导转换突变。这种突变可以第二次用2AP处理而发生回复突变。依据其最初整合

到 DNA 中的状态和它在 DNA 复制时的状态，突变包括 AT 到 GC 的转换或 GC 到 AT 的转换。由于 2AP 和 5BU 诱发的突变都是转换，所以 5BU 可以逆转 2AP 诱发的突变或者相反。

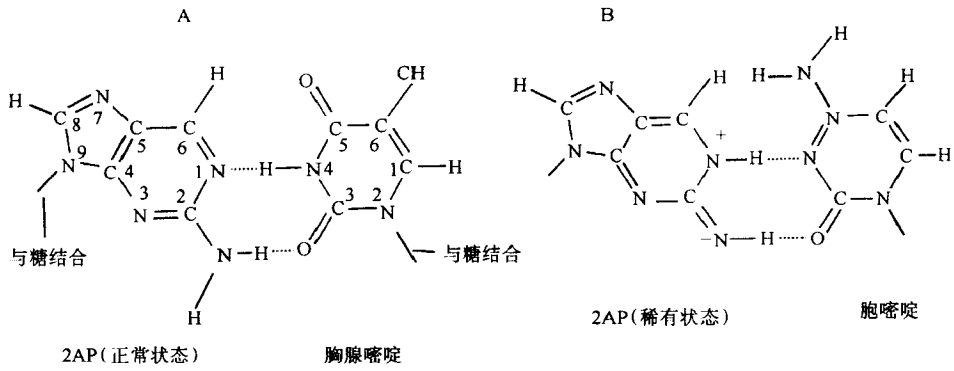


图 2－15　碱基类似物 2－氨基嘌呤（2AP）的碱基配对特性

A. 正常状态的 2AP 与 T 配对　B. 稀有状态的 2AP 与 C 配对

2. 碱基修饰剂　同碱基类似物相似，DNA 在复制过程中吸收碱基修饰剂。这类化合物通过修饰碱基的结构和特征来诱发突变。图 2－16 示三种突变剂以这种方式诱发突变的过程。这三类突变剂为脱氨基因子（deaminating agent）、羟基化因子（hydroxylating agent）和烷化剂（alkylating agent）。

亚硝酸为脱氨基因子，它可以从鸟嘌呤、胞嘧啶和腺嘌呤碱基上除掉氨基（$-NH_2$）。用亚硝酸处理鸟嘌呤产生黄嘌呤，但由于这种嘌呤碱基的配对特性同鸟嘌呤相同，所以只会造成中性突变。但胞嘧啶经亚硝酸处理后则产生尿嘧啶。由亚硝酸诱发的胞嘧啶的脱氨基作用在 DNA 复制期间产生 CG 到 TA 的转换突变。同样亚硝酸修饰腺嘌呤而产生次黄嘌呤（hypoxanthin）。次黄嘌呤与胞嘧啶配对而不与胸腺嘧啶配对，从而导致 AT 到 GC 的转换突变。由亚硝酸诱发的突变经第二次用亚硝酸处理可使其发生回复突变。

另一种碱基修饰突变剂为羟氨（NH_2OH）（hydroxylamine）。它与胞嘧啶进行特殊反应，通过加上羟基（OH）而对其加以修饰。羟基化的胞嘧啶只能与腺嘌呤配对而不与鸟嘌呤配对。因此由羟氨诱发的突变为 CG 到 TA 的转换。由于羟氨只是以一种方式诱发突变，因此这种突变不能第二次用这种化合物处理而发生回复。但用其他突变剂如 5BU、2AP 和亚硝酸处理可使其发生回复突变，即 TA 到 CG 的转换突变。

最后一种碱基修饰剂为硫酸甲基甲烷（methylmethane sulfonate，MMS）

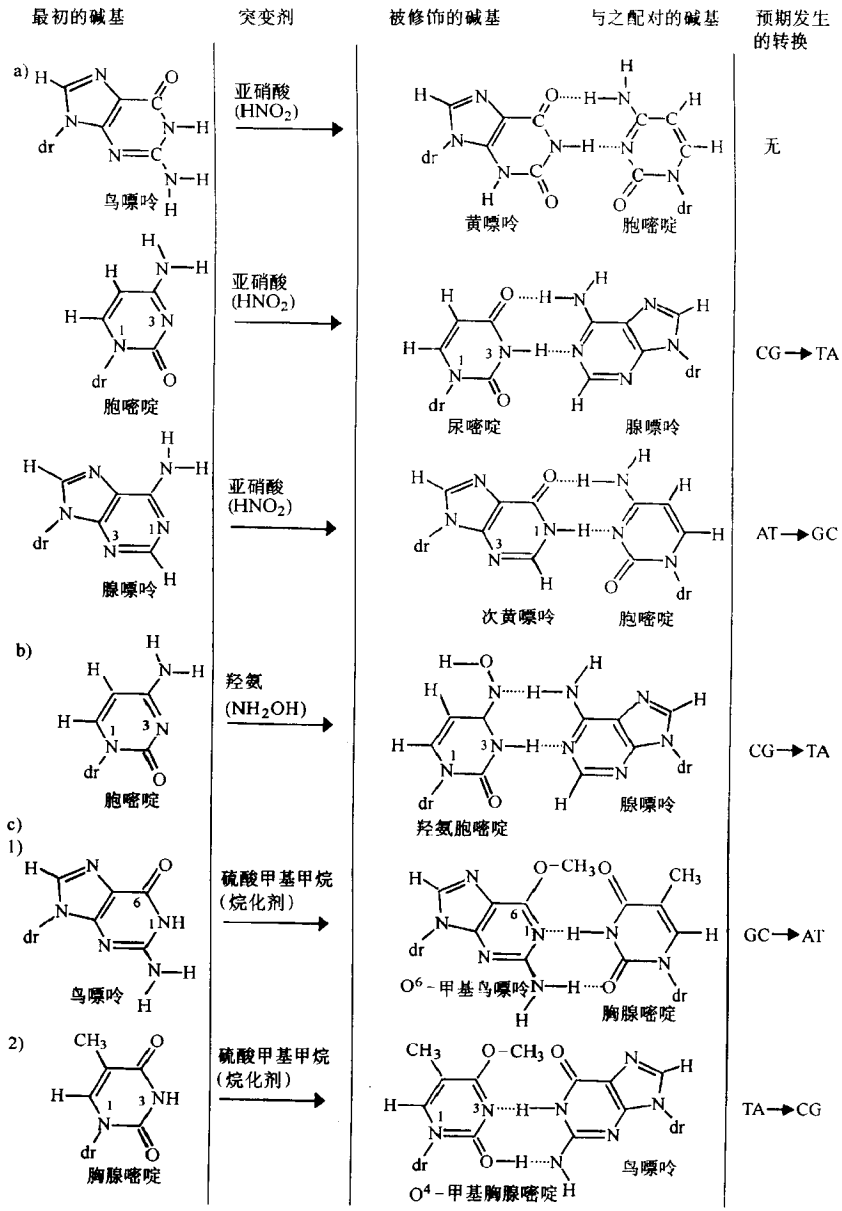


图 2-16　三种碱基修饰剂的作用

a. 亚硝酸修饰鸟嘌呤、胞嘧啶和腺嘌呤，胞嘧啶和腺嘌呤被修饰后导至突变，而鸟嘌呤被修饰后则不会引起突变　b. 羟氨只与胞嘧啶反应　c. 硫酸甲基甲烷为一种烷化剂，烷化鸟嘌呤和胸腺嘧啶

这种诱变剂将烷基（如 $-CH_3$ 和 $-CH_2CH_3$）加到碱基的各种位置上。烷化剂将一个碱基进行修改，使其与另一错误碱基配对。如图 2－16 所示，经 MMS 处理后，一个甲基化的鸟嘌呤与胸腺嘧啶配对，发生 GC 到 AT 的转换，而甲基化的胸腺嘧啶则与鸟嘌呤配对，发生 TA 到 CG 的转换。

3．嵌入剂　嵌入突变剂（intercalating agent）包括原黄素（proflavin）、吖啶（acridine）、溴化乙锭（ethidiumbromide）和 ICR－170 等。它们可插入到 DNA 双螺旋的一条链或两条链的相邻碱基之间。

如果嵌入剂插入到一条 DNA 单链的两个相邻碱基之间，嵌入剂相对的一条互补链中可能会随机地加入一个碱基（图 2－17），当经过一至数轮 DNA 复制之后，嵌入剂随之丢失。由于插入了一个碱基对，最后就会导致移码突变。如果嵌入剂插入到新合成链中并占某个碱基的位置，当嵌入剂丢失后，DNA 双螺旋经过复制，也会因缺失一对碱基而导致移码突变。

移码突变的结果是突变位点以下所有氨基酸都可能是错误的，最后形成的蛋白质可能不具有功能。由嵌入剂诱导的移码突变再用这些嵌入剂处理可以发生回复突变。

（五）离体条件下 DNA 分子的位点专一性诱变

自发突变和诱发突变都是随机地发生在基因组的任何位点上，而不能限制在某个特殊基因之内。遗传学家希望研究的是某些特定基因中发生突变的效应。利用辐射或化学突变剂处理细胞或生物体之后，必须对生存的群体进行筛选，以鉴定目标突变体。采用 DNA 重组技术，有可能克隆到目标基因和合成大量用于分析和操作的 DNA。也就是说现在可以在试管内修改 DNA，并可通过感染或转化（见第九章），再将其送回到细胞之中，最后测定突变的效应。这些技术可以使遗传学家在某个特殊基因内的特定核苷酸位置创造突变。这种程序叫做 DNA 的位点专一性诱变（site－specific mutagenesis）。

位点特异性突变可用于创造点突变或小片段的缺失和插入。图 2－18 示创造点突变的过程。

首先是制备某个基因或部分基因片段的克隆，DNA 经变性后，再将一段寡核苷酸（一段短的、合成的 DNA 片段）联会到其中一条单链上。寡核苷酸是人工合成的，它可与克隆的 DNA 的部分区域进行互补。在本例中，我们的目标是将 GC 碱基对变成 AT 碱基对，因此寡核苷酸具有 TAGACTAT 序列，划线的 A 为突变位点。在克隆的 DNA 的一条单链上与其相对应的位置有一个 C。一旦寡核苷酸联会上以后，DNA 聚合酶就以它为引物，合成一段完全互补的环状模板链。然后 DNA 连接酶再催化新链形成环状分子。此时我们就得到

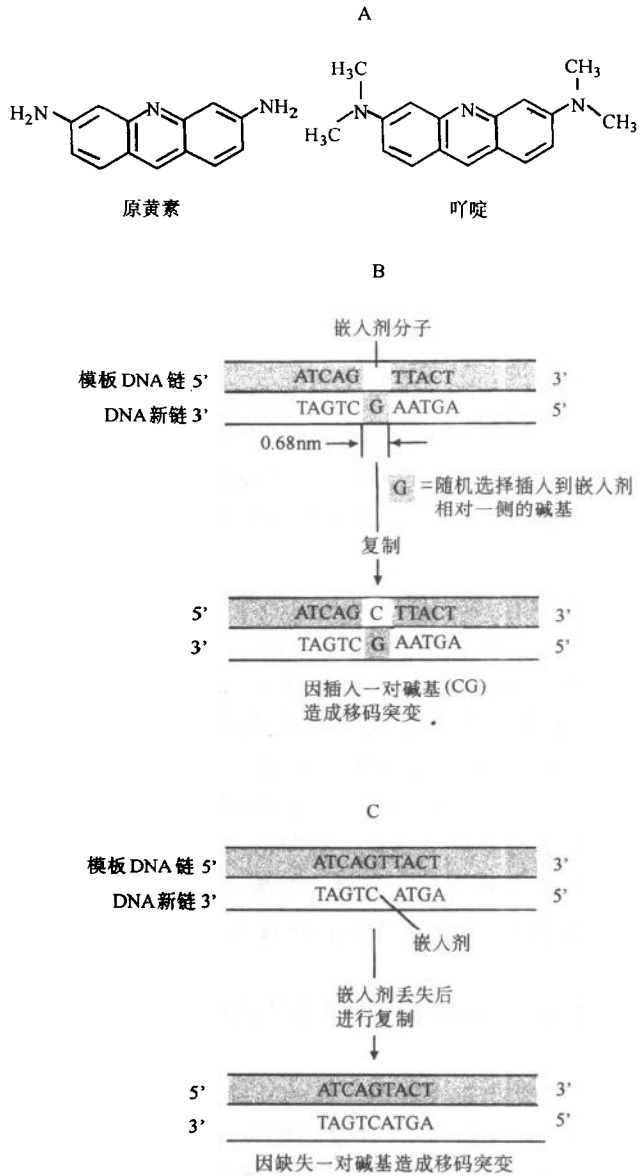


图 2－17　嵌入剂造成移码突变的机制

A. 代表嵌入剂的结构　B. 插入突变　C. 缺失突变

一条在一个位点上有误配对的 AC 碱基对的环状双链 DNA 分子。然后再通过转化，将这种 DNA 引入细胞中。该 DNA 通过复制就可产生两种分子，一条

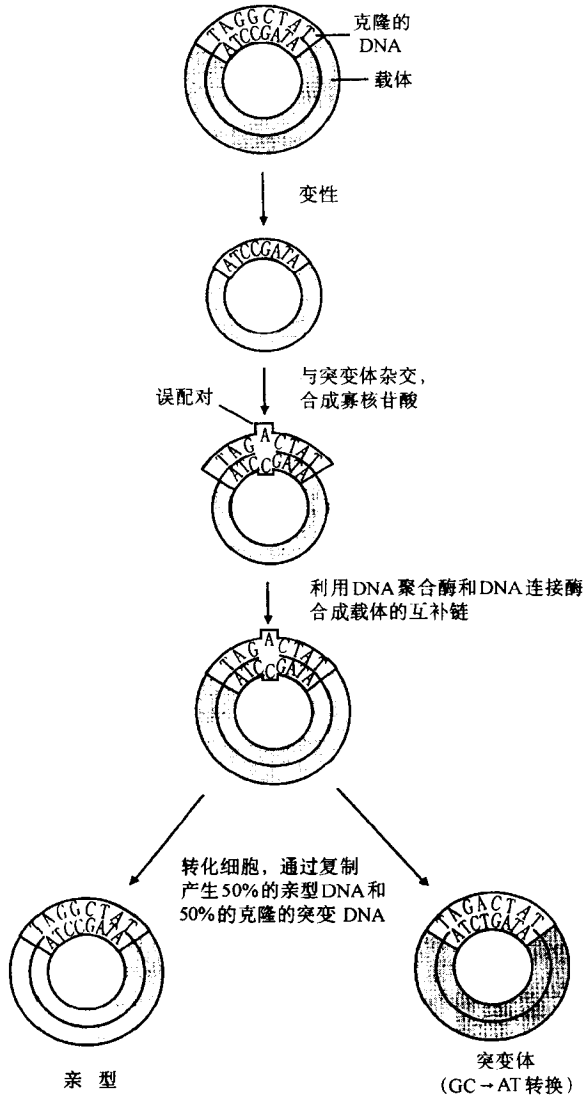


图 2－18　在特定位点上创造点突变的过程

是在目标位点上具有 GC 碱基对的重组分子，另一条是在目标位点上具有 AT 碱基对的突变分子。采用这种方法时，如果利用比互补模板链少一个或多一个碱基的寡核苷酸，我们就可创造单个碱基对插入或缺失的突变分子。

采用这种技术，我们可以对一个未知功能的基因进行突变，然后再将突变的基因引入细胞中，研究其表型变化，从而了解正常基因的功能。这类实验

研究叫做逆遗传学（reverse genetics）。也就是说，传统的遗传学是从表型开始，向回追溯控制该种表型的基因，而逆遗传学则是从基因开始向前追踪该基因编码的功能。位点专一性诱变还应用于在某些关键的调节序列中创造某种特殊变异，以研究哪些碱基对是重要的，哪些碱基对是不重要的。我们对启动子或增强子的认识有一部分就是从对基因上游或下游的序列进行位点专一性诱变而获得的。

二、基因重组

现在普遍认为，DNA 分子的重组（recombination）并不是一种偶然现象，而是细胞内一种由酶催化的基本过程。这类酶除了为细胞提供遗传变异以外，而且还能使细胞重新获得丢失的 DNA 序列。此外，某些特殊类型的重组还调节基因表达，细胞通过对其染色体中某些特殊的片段进行转换，将显性基因置入能够表达的位点上，甚至创造新的蛋白质编码区。

（一）交换的断裂重接模型

根据对减数分裂过程的细胞学观察，交换是由于配对的、相互缠绕的一对同源染色体因收缩所产生的张力而使染色单体折断的结果。折断的末端通过交替重接来缓解这种张力，从而产生两条互换了片段的重组染色体。依据这种模型，重组过程发生在染色体复制之后，也就是说，重组发生时，每一对联会的染色体都有 4 条染色单体。后来发现，这种断裂重接现象不仅可以发生在同源染色体的非姊妹染色单体之间，而且还可以发生在姊妹染色单体之间。因此，交换也是姊妹染色单体之间进行 DNA 互换的过程。

根据上述理论，很显然，在交换过程中 DNA 分子必须准确联会在一起，因此交换必然发生在两条相同或近乎相同的 DNA 序列之间，因为两条不同染色体之间通常不可能进行碱基配对而使二者联会在一起。虽然交换通常发生在同源染色体的相对应的区域，但也可发生在非对应区域的同源部分之间，其结果就是造成不等交换（unequal crossing over）。不等交换往往会造成染色体重复和缺失。

虽然有关染色体断裂—重接的分子基础目前还不清楚，但是大量实验证明，两条 DNA 分子的断头之间确实可以重新结合在一起。根据 DNA 分子的物理特性，两条 DNA 分子是不可能简单地进行断裂重接的，其中可能涉及若干种催化重组过程的酶类。

（二）普遍性重组的霍利迪模型

高等真核细胞在减数分裂过程中进行染色体互换的机制除了上述断裂重接

模型以外，目前还存在其他一些解释这种普遍性重组的模型，其中最为熟知的模型之一为霍利迪模型（Holliday model）。这种模型首先由 R.Holliday（1964年）提出，后来又得到许多人（包括霍利迪本人）的修改。尽管如此，在重组的许多具体步骤方面，霍利迪模型可能并不完全正确。

根据霍利迪模型，重组的第一个步骤是两条双螺旋染色体中同源核酸序列的识别和联会（图 2－19）。接着，在两条具有相同极性 DNA 单链的相对应的位置形成缺口。然后断裂链侵入相对的双螺旋中，进行碱基互补配对。最后连接酶缝合缺刻，形成一种类似希腊字母 X 的结构，这种结构叫做霍利迪中间体（Hollidoy intermediate）或霍利迪连接体（Holliday junction），有时也叫半染色单体交叉（half chromatid chiasma）。

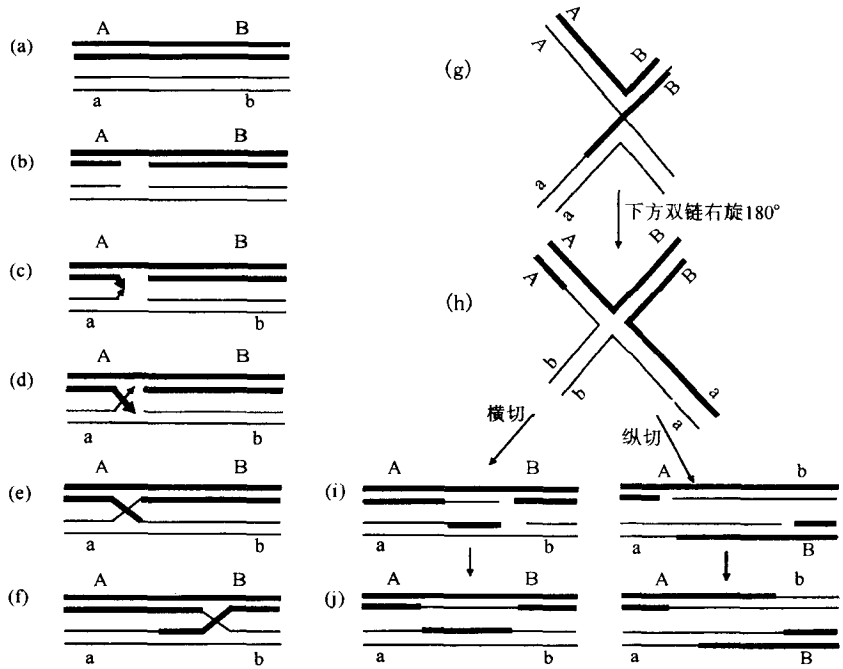


图 2－19　重组的霍利迪模型

示 DNA 片段互换

霍利迪结构并不稳定，当两条双螺旋以相同方向转动时，这种交叉连接可能向左或向右移动，形成一段杂双链区，这一过程叫做交连桥前移（bridge migration）。如果霍利迪结构的四个末端在一个平面上分开，就可形成图 2－19g 的构型。如图 2－19g 中下方一条双螺旋沿其交叉链转动 180°，就形成图 2－19h的结构。这种结构相当于电镜下观察到的 DNA 重组中间产物的实际构

型。如果霍利迪结构在交连桥区域横切或纵切，就产生图 2－19i 中 4 种可能的产物。在横切的两种产物中，两条双螺旋中的标记基因 AB 和 ab 都保持了原来亲链状态，而纵切则产生了 Ab 和 aB 两种重组子。但是，无论纵切或横切，在 A 和 B 座位之间都有一段杂双链区，也就是说这一区域中两段单链序列分别来自两条不同的 DNA 分子。在图 2－19i 中，每一条 DNA 分子都含有一个缺口或缺刻，这些缺口最后都由 DNA 聚合酶和 DNA 连接酶修补，形成完整的 DNA 双螺旋（图 2－19j）。

（三）由 DNA 损坏启始的重组

虽然上述霍利迪模型可以说明重组期间某些关键步骤，但各种生物并不只是通过惟一途径形成霍利迪中间产物。根据对各种细菌的研究，大多数同源重组往往都由单链缺口或缺刻（包括双链断裂）所引起，从单链缺口和缺刻开始，通过不同途径形成霍利迪连结体（图 2－20）。引起 DNA 分子中形成缺口的因素很多，某些外部因素如紫外线、X 射线以及一些化学诱变剂等都能诱导双链 DNA 分子形成单链缺口，所以这些缺口或缺刻实际上都是 DNA 分子中的损坏斑，因此这种在 DNA 分子的损坏位点上进行的重组实际上就是一种 DNA 修复过程如重组修复（见第一章第四节）。

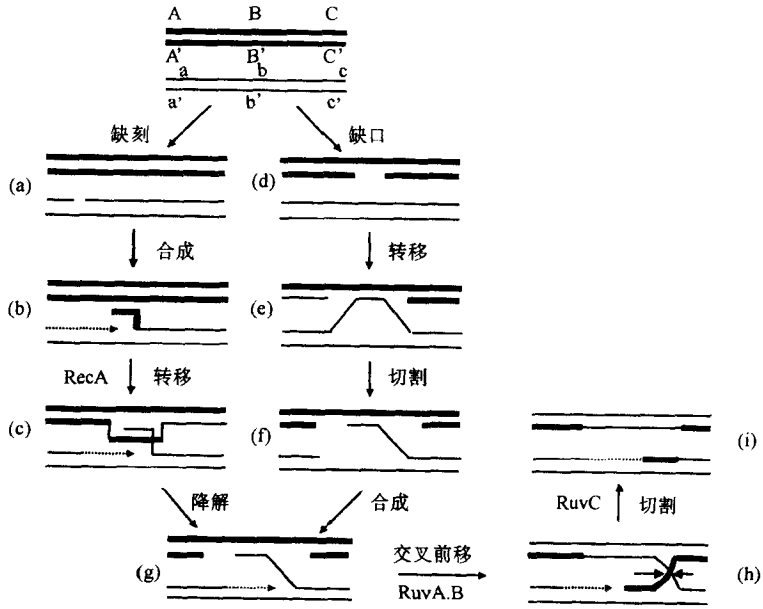


图 2－20　单链缺口启始形成霍利迪中间产物的途径
虚线为修复合成，小箭头为切割位点

这种修改的重组模型有两个重要特征，第一，它须进行单链转移，或者是将含有缺刻的双螺旋的一条单链转移到完整双螺旋的同源区中（图 2－20c），或者是完整双螺旋的某一段单链转移到具有互补序列的单链缺口上（图 2－20e），从而使受体链都可能含有碱基不完全互补的杂双链区。第二是要进行修复合成，即单链转移后留下的缺口须由 DNA 聚合酶进行修复（图 2－20b 和 g）。当两条双螺旋形成交叉连接之后，如果交叉不前移，切割交叉之后，其中有一条双螺旋含有杂双链区，另一条则保持原来的状态。如果交叉前移，则两条双螺旋都含有杂双链区。这些杂双链区以后通过误配对修复或 DNA 复制，都可能转变成纯双螺旋。此外，在这个模型中，重组中间产物都只按一种方式切割，即只切割交换的单链，因此，这可能是细菌进行同源重组的一种主要方式。例如，某种线状分子同细菌环状染色体重组后，这种重组方式就不会使环状分子变成线状分子。如果按前述霍利迪重组模型或断裂重接模型，必须是偶数重组才能使细菌环状染色体保持原状。

（四）细菌中参与重组的蛋白质

根据对噬菌体、细菌和酵母的研究，这些生物都具有催化重组过程的酶，如大肠杆菌中，由 recA 基因编码的产物就是一种重要的重组酶。RecA 蛋白是一种具有多功能的酶，在重组过程中的基本作用是促进 DNA 单链转移。根据对大肠杆菌结合过程中同源重组的研究，RecA 蛋白能够使转移到受体细胞中的单链 DNA 结合到同源的双链 DNA 上，形成一种 D 环（图 2－21）。在这个过程中，RecA 蛋白首先与单链 DNA 结合，并使这条单链与同源 DNA 双链中的互补序列进行碱基配对，以替换双链中一条单链。在存在 ATP 的条件下，RecA 蛋白可能通过将双螺旋的不同部分解螺旋，然后再将与其结合的单链 DNA 与解螺旋的区域联会在一起，并使转移的单链按 5’→3’方向与双链中的互补链逐步进行碱基配对。当杂双链区形成后，RecA 蛋白脱离复合体。

各种原核细胞中都普遍存在 RecA 的同系蛋白，而且在真核细胞中也存在与其相似的蛋白质。例如在啤酒酵母中有 2 种蛋白质与 RecA 很相似，二者都参与染色体的联会过程。可能真核细胞中染色体联会与原核细胞的 DNA 单链转移反应很相似。

虽然 RecA 蛋白在重组过程中执行基本的碱基配对功能，但在细菌中还有其他蛋白质与重组过程的另一些步骤有关，其中较为熟悉的就是 RecBCD 酶。RecBCD 酶又称外切核酸酶 V，为一种复合体，分子量约为 300kd，也是一种多功能蛋白，具有使双链 DNA 解螺旋的作用，同时对单链 DNA 分子也具有外切核酶和内切核酸酶活性。此外，它还具有一种依赖于 DNA 的 ATP 酶活性。RecBCD 酶有一个重要特征，即它只从双链 DNA 分子的齐头末端启始解

螺旋过程。当 RecBCD 酶结合到双链 DNA 末端之后，他便利用 ATP 作为能量，沿双螺旋前移，一边将 DNA 解螺旋，一边使 DNA 再度螺旋化。由于 RecBCD 酶将 DNA 解螺旋的速度比使两条单链形成双螺旋的速度快，所以解螺旋的片段不断增长。

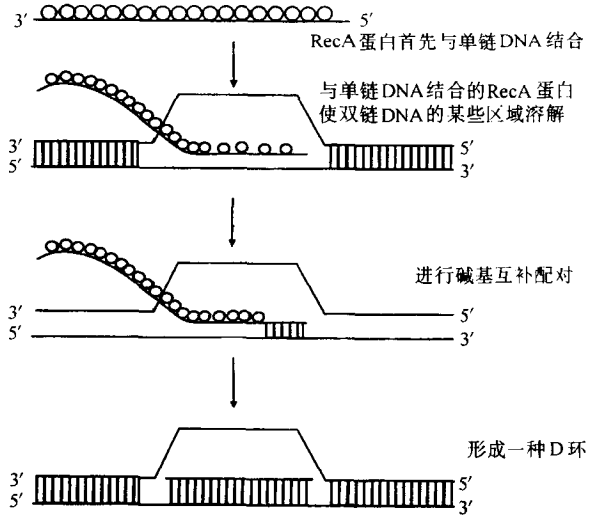


图 2－21　RecA 蛋白使转移到受体细胞中的单链 DNA 结合到同源的双链 DNA 上，形成 D 环

在 DNA 解螺旋过程中，RecBCD 酶的第二种功能是在叫做 X（chi）位点（X site）的区域促进重组。X 位点的序列为 GCTGGTGG。当解螺旋的 DNA 分子中存在 X 位点时，RecBCD 酶切割 X 位点附近暴露在外的单链，因此当其继续前移时，就能阻止其后的 DNA 再度螺旋化（图 2－22），留下游离的单链末端和单链缺口。于是 RecA 蛋白就能与这种单链尾巴和缺口结合，启始 DNA 链与另一同源序列进行交换。

据统计，大肠杆菌染色体上约有上千个 X 位点，约每 5 个基因中就有一个，可以为 RecBCD 酶的作用提供大量机会。在正常的大肠杆菌细胞中，这些序列通常不能作为 RecBCD 酶的作用位点，因为其 DNA 不具有游离末端。但是在结合过程中，由于来自供体细胞的 DNA 具有一个游离的 DNA 末端，RecBCD 酶就能与 DNA 末端结合，当运动到 X 位点时，就可切割单链，促进其与受体细胞 DNA 进行重组。在受 λ 噬菌体感染的细胞中，DNA 也存在游离的末端，但这种末端是在噬菌体 DNA 的包装过程中，由 λ 噬菌体的酶进行切割而形成的。

大肠杆菌中还有一组由三个基因编码的蛋白质参与重组的后期过程，即

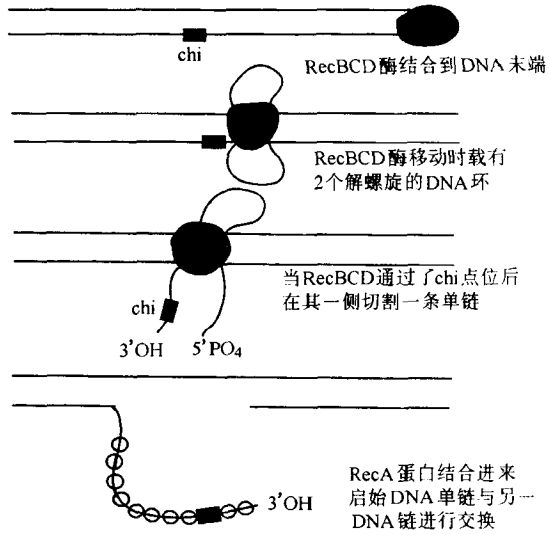


图 2－22　RecBCD 酶参与重组的模型

ruvA、B、C 三个基因的编码产物。RuvA 和 RuvB 加强杂双螺旋结构的形成，其中 RuvA 识别霍利迪连结体，RuvB 为一种 ATP 酶，为交叉前移提供能量，所以 RuvAB 复合体的作用是使交叉前移。RuvC 为一种内切核酸酶，能特异性地识别和切割霍利迪连结体。

即使在大肠杆菌中，我们也只知道某些基因或酶类与重组过程有关。在大肠杆菌中，除了 recA、B、C、D 基因和 ruvA、B、C 基因以外，还有其他一些基因的编码产物与重组过程密切相关。有些与 DNA 合成有关的蛋白质也参与重组过程，如单链 DNA 结合蛋白（第三章）具有促进 RecA 蛋白与单链 DNA 结合的功能。此外，重组过程还需要 DNA 聚合酶修补缺口，同时也需要 DNA 连接酶将重组分子的末端连接在一起。

三、基因转变

如上所述，交换位点上的重组过程并不始终都是相互的，在交换的过程中，往往伴随另一种重组现象即基因转变（gene conversion）。基因转变是指两个基因在互作过程中，其中一个基因的核苷酸序列整合到另一基因上，整合后这两个基因都保持原有位置和完整程度，但其中有一个基因在结构上不发生互换的现象。

根据对子囊菌（*Ascomycetes*）中酵母（*Saccharomyces*）、链孢霉（*Neurospora*）、粪

生粪壳菌(*Sordaria*)等的四分子分析或八分子分析，基因转变发生后，其子囊孢子在子囊内的分离则偏离4+:4-的正常比例，而出现6+:2-、2+:6-、5+:3-、3+:5-、1+:7-、7+:1-或异常4+ :4-分离等，即所谓减数后分离（post meiotic segregation)。在高等真核生物中，基因转换发生后，配了的分离比例则可能呈3:1或1:3分离，而不是呈2:2正常分离。

前述各种重组模型都能说明基因转变发生的机制。例如，在霍利迪模型中，无论中间产物进行横切或纵切，两条双螺旋都含有一段杂双链区，这些杂双链区通过修正，如果其中一条单链回到供体状态或受体状态，子囊菌中子囊孢子则呈5+:3-或3+:5-分离，如果一段杂双链的一条单链在校正后恢复到受体,另一段杂双链的一条单链则恢复到供体状态,子囊孢子就会呈6+:2-或6-:2+分离（图2-23)。如果杂双链区不进行校正,经过减数分裂后的DNA复制之后，则子囊孢子呈异常4+:4-分离。

除了前述有关重组模型以外，还有其他一些模型也能解释基因转变发生的机制，其中之一是双链断裂修复模型（double strand break repair model)。双链断裂修复假说认为，基因转变是由DNA双链上的切口所引起，如果切口再扩大就会形成较大的缺口，在DNA复制过程中，细胞通过修复DNA双链缺口而实现基因转变（图2-24)。根据这种模型，子囊菌中子囊孢子呈6:2（或2:6)分离。但是，在这个模型中，由于双链修复区的两端都接有杂双链区，依据杂双链区中校正过程的差别，也会出现5:3（或3:5）和6:2等减数后分离比列。

双链断裂修复模型目前已有可靠事实作依据，根据对啤酒酵母(*Saccharomyces cerevisiae*）精氨琥珀酸裂解酶（argininossuccinase）基因Arg4的研究，基因转变发生时有一个转变启始位点（conversion initiation site)，即形成双链缺口的位置。这个启始位点是相对固定的，在Arg4基因中，它位于该基因启动子之中，即在转录起点上游-300到-30的一段序列之中。由于基因转变受到启动子的制约，因此属于同一复等位基因系列中的不同突变等位基因就具有不同基因转变频率。具体地说，也就是在一个基因座位中，5'位点转变频率较高，3'位点转变频率较低。例如，在Arg4座位上有7个不同的突变体Nsph、Acc、Rv、Bcl、Dra、Aha和Bgl，当这些位点上发生突变后，分别影响限制性内切酶NsphⅠ、AccⅠ、EcoRⅤ、Bcc I、DraⅢ、AhaⅡ和BglⅡ的切割反应，这些突变体的杂合二倍体（m/+）之间基因转变频率各不相同如图2-25，其中5'端Nsph位点转变频率最高，3'端Bgl位点转变频率最低。这种在某个基因座位内，各个突变位点上基因转变的频率从基因座位的5'端到3'端呈现一种梯度降低的现象叫做转变极性（ conversion polarity)。

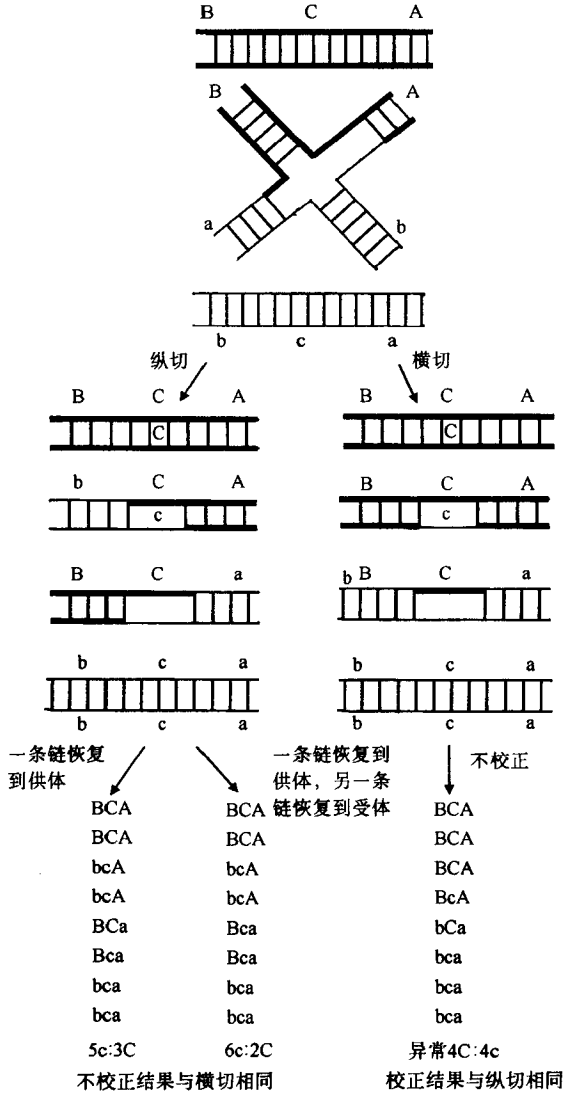


图 2－23　霍利迪重组模型中基因转变后子囊菌子囊孢子的各种分离类型

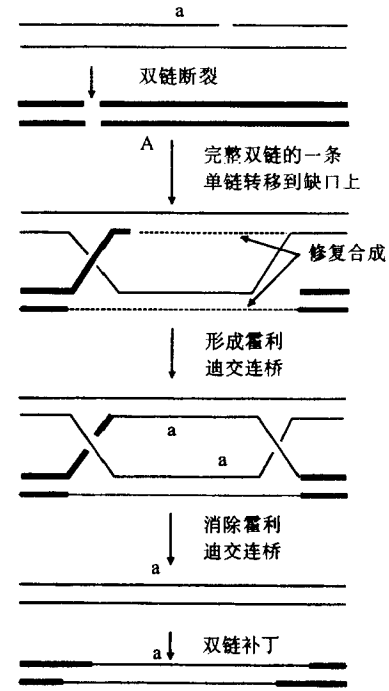


图 2－24　基因转换的双链断裂修复模型

（只绘出其中 2 条染色体）

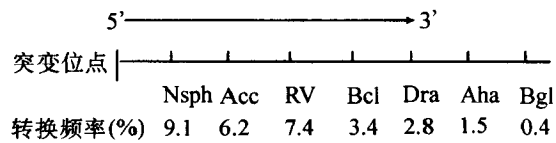


图 2－25　酵母 Arg4 座位上不同突变等位基因的转换极性

长箭头示转录方向

造成子囊菌中减数后分离的机制可能涉及多种重组类型，除了前述各种重组模型以外，基因转变也可能是DNA分子在复制过程中使用了错误模板，如不同染色体、同一染色体的不同区域甚至RNA的结果。

第四节　转座子

在原核生物和真核生物基因组中，许多重复序列可以从基因组的某一位置自由地转移到另一位置，这类DNA序列叫做转座子（transposon）、转座因子（transposable element）、插入序列（insertional sequence）或跳跃基因（jumping gene）。转座子切割和转座后，往往造成基因突变或染色体畸变。本节主要以细菌、植物和果蝇基因组中几种转座因子为例，对转座子的遗传特征及其对宿主基因组遗传变异的影响加以讨论。

一、转座子的遗传特征

转座子是DNA分子的一条片段，有些相当短，有些相当长。转座子以相当高的频率从基因组的一个位点转移到另一位点上，当插入到某个基因座位以后，会影响该基因及其邻近基因的表达。由于转座子在DNA分子中的位置不稳定，因此转座后对插入位点基因的影响也可能随之消除。通常，转座子转座时所需的酶类由转座子本身所编码。

转座子的主要特征如下：

（1）当转座子插入到某个基因之中或附近时，往往会降低或完全抑制这一基因表达，造成所谓的隐性突变（recessive mutation）。转座子转座后会使该基因恢复正常功能，但因转座子序列的结构特征不同，基因功能的恢复程度也有所不同。

（2）转座子转座后，在转座子两端产生两段短的重复序列，其中左端的一段重复序列是受体DNA位点上原有的一段序列，而右端的重复只是左端重复的一个拷贝，其长度和序列同源性因转座子不同而有别。

（3）大多数转座子在右端都有AATAAA结构，这段序列与mRNA的多聚A化有关。

（4）在许多情况下，转座子的一部分序列可以转录mRNA，其编码产物有些与转座有关。

（5）在同一生物的基因组中可能存在数种不同的转座子。

（6）有些转座子为一种复合结构，由2个或2个以上不同转座子组成，每

一成员都可能影响邻近其插入位点的基因的活性。

二、细菌中主要的转座子

在细菌中，转座子主要有两种类型，即 IS（insertional sequence）和 Tn（transposon）家族，每一类型又包括许多种不同成员。

（一）Tn 因子

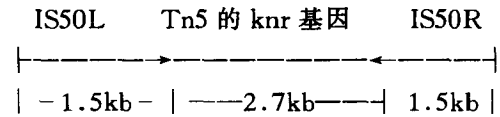
在利用细菌的抗药性质粒（R 质粒）构建载有卡拉霉素（kanamycin）抗性基因的 λ 噬菌体的过程中，发现抗药性基因的传播非常快，它可以从一种 R 质粒跳跃到另一质粒上，也可以跳跃到另一些 DNA 序列如噬菌体或细菌染色体上。根据 DNA 分子杂交知道，与卡拉霉素抗性有关的 DNA 序列长 5.2kb，两端接有长 1.5kb 的颠倒重复序列，当抗药性基因跳跃时，颠倒重复序列也随之转移。具有这种结构和特性的 DNA 序列就叫做转座子或 Tn 因子（表 2-2）。上述这种 Tn 因子叫 Tn5，它由 Tn 本身和两段叫做 IS50 的插入序列即 1.5kb 的颠倒重复序列组成。有关 Tn 因子的特征现以 Tn5 和 Tn3 为例加以说明。

表 2-2 部分 Tn 因子的特征

转座子	所载基因	大小（kb）	IS 因子	重复序列
Tn1	Amp^R	5	—	—
Tn3	Amp^R	4.957	—	颠倒重复
Tn4	Amp^R	20.5	—	颠倒重复
	汞离子抗性		IS50	
Tn5	Kan^R	5.7	IS50	颠倒重复
Tn6	Kan^R	9.2	—	—
Tn7	$Trim^R$	13.5	—	颠倒重复
Tn9	Chl^R	2.5	IS1	顺式重复
Tn10	$Tetr^R$	9.3	IS10	颠倒重复

注：Amp 氨苄青霉素，Kan 卡拉霉素，Trim 三甲氧苄二氨嘧啶，Chl 氯霉素，Tetr 四环素。

Tn5 属于一类复合转座子，结构比较特殊，其主序列长 2.7kb，位于中央，两个 IS 序列以相反的极性位于主序列两侧如下图。主序列含有几个抗生



素抗性基因，包括卡拉霉素、链霉素（streptomycin）和博莱霉素（bleomycin）一类的抗性基因。Tn5 主序列两则的 IS 因子既是 Tn5 的重要组成部分，同时也是一种自主性的转座子（图 2－26）。位于 Tn5 右侧的 IS50R 编码两种与转座有关的酶，其中一种（蛋白 1）为转座酶，另一种（蛋白 2）为转座抑制蛋白，这两种蛋白质都由同一段 DNA 序列编码。Tn5 左侧的 IS50L 与 IS50R 只有一对碱基的差别，但这一微小差别会产生两种重要影响，（1）在转座酶编码区内造成琥珀突变（ochre mutation），产生翻译的终止信号，所以这种突变蛋白不具有转座功能；（2）形成一个 Tn5 主区中卡拉霉素抗性基因的启动子（P2），所以 IS50L 对主 Tn 因子的转录具有重要影响。Tn5 的这种结构特征表明，两个 IS 夹住一个 Tn5 主序列并促使其转座到新的位点上。Tn5 转座时，首先从染色体上切割下来，形成环状结构，然后再插入到新的座位上。因此，Tn5 只是从一个位点转移到另一位点，在原位并不留下一个拷贝。Tn5 转座到新的位点上后，两端都形成一段顺式重复序列。

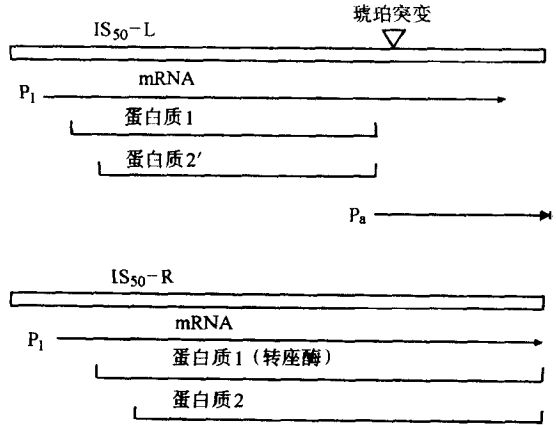


图 2－26　Tn5 的 IS50－L 和 IS50－R 的组成

Tn3 转座子不含 IS 序列，只是在两端接有短的颠倒重复（图 2－27）。Tn3 长 4 957bp，其中有 3 个编码区，最大的一个叫做 tnpA，编码转座酶（1 015 个核苷酸），另外两个顺反子叫做 tnpR 和 bla，分别编码一种叫做溶解酶（resolvase，长 185 个氨基酸）和一种叫做 β－酰胺酶（lactamase，长 286 个氨基酸）的蛋白质。后两个顺反子的转录方向与 tnpA 基因相反。

Tn3 有三个转录启始位点，第一个位于 tnpA 的 5’方向，第二个位于转录方向相反的 tnpA 和 tnpR 基因之间，第三个位于 tnpR 的 3’端和 bla 编码区的前导序列之间，可能它们分别为 tnpA、tnpR 和 bla 三个基因的转录启始信

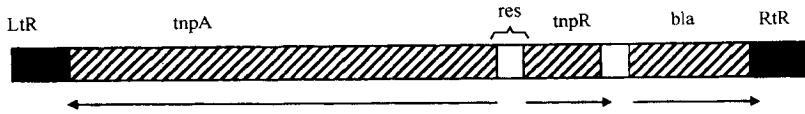


图 2－27　Tn3 的结构

Ltr 和 Rtr 分别为左右两个末端颠倒重复

号。在这三个位点之中，转录酶的识别序列分别为 TATAATA，CATAATA 和 GACAATA，其相应辅助位点的结构可能分别为 AACGAAG、GT CCATT 和 ATTCAAA。

Tn3 转座需要两种以上的酶和一对末端重复序列，同时也需要转座子中部一个叫做 res 的位点，如果缺少这一位点，转座子就不能完成转座过程，而是将其一个拷贝转移到新的位点上，并使新、老位点相连，其结果就使供体和受体复制子融合。融合体可以由溶解酶分割成两个复制子，因为溶解酶具有抑制转座酶合成和催化融合体解体的功能；融合体也可通过两个 res 位点之间重组而分开成供体和受体两个复制子（见本节后）。

（二）IS 因子

根据前述 Tn5 的结构知道，IS 因子通常与其他 Tn 因子结合，但是在 Tn 家族中，IS 则是一种完全不同的转座子，其主要差别在于 IS 不具有编码抗生素抗性的基因。细菌中绝大多数 IS 序列都具有两种特殊结构。第一，IS 两端都有一段 10～40bp 的颠倒重复序列。第二，IS 序列具有一至二个蛋白质编码区，在编码区上游存在启动子、核糖体结合位点、ATG（AUG）起点密码；编码区下游接有肽链终止密码。

大肠杆菌中另一种转座因子 IS1，长 768bp，位于 Tn9 和 Tn7 的两侧（图 2－28）。IS1 共有 6 个可读框，其中至少有两个基因是转座过程不可缺少的，即 insA 和 insB 基因，二者结构相似，极性相同。IS1 两端的颠倒重复序列也是转座不可缺少的，它可能是转座的识别信号。另外，每一颠倒重复序列都含有一个启动子，控制从 IS1 末端开始向内部转录的过程。

IS1 转座后，在其末端形成两段染色体顺式重复，在不同插入位点中，其长度不等。现在知道，顺式重复的长度变化是 IS1 自身的一种特征，野生型和突变型 IS1 都能产生 8bp 或 9bp 的顺式重复，典型的顺式重复长 9bp。通常顺式重复富含 A－T 碱基对，但某些情况下也富含 G－C 碱基对。

IS1 两端的颠倒重复在结构上能够互补，但其中也存在一些不能配对的碱基。在 IS1 的一种突变体 IS1－16 中，其左端的颠倒重复含有启始 insA 和 insB 基因转录的启动子，而右端的重复可能与一种所谓的拮抗信使(antimessenger)合成有关，这种拮抗信使对左端启动子具有负控制功能，但这种假说还有待证实。

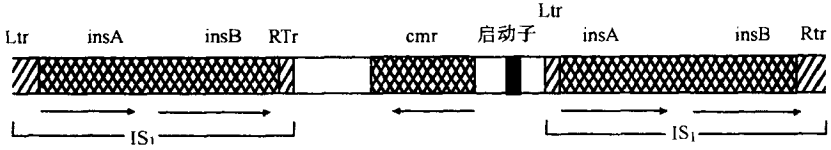


图 2－28　Tn9 中的 IS 因子

Ltr 和 Rtr 分别为左、右末端重复、cmr 为氯霉素抗性基因

（三）Mu 噬菌体为一种巨型转座子

噬菌体是一类细菌病毒，而 Mu 噬菌体（Mu phage）则是一种巨型转座子。第一，Mu 噬菌体整合到细菌染色体上后造成插入位点上染色体 DNA 发生 5 对碱基重复。第二，Mu 噬菌体并不从染色体上切割下来，也不在细胞质中复制。虽然受感染的细胞在裂解后可产生数百个 Mu 噬菌体，但在裂解周期中，从未发现游离的 Mu DNA 分子，Mu DNA 只有通过转座才能复制，并且直接在寄主 DNA 的整合位点上进行包裹。

Mu 噬菌体在宿主染色体上对其 DNA 进行包裹的过程中，其头部结构首先识别噬菌体左端的、从第 50 至 100bp 之间的一段 DNA，然后再开始包裹。包裹进来的 DNA 包括 Mu DNA 两端的染色体 DNA，当其中一部分染色体 DNA 被切除掉以后，才将尾部接上。被包裹的 DNA 比噬菌体基因组稍大，包括约 3kb 的细菌 DNA，其中主要为 Mu DNA 右端的染色体 DNA。

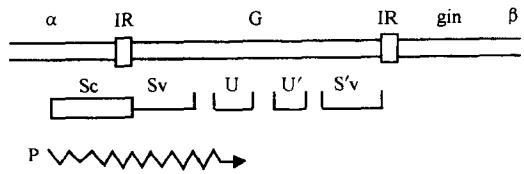


图 2－29　Mu 噬菌体 G 区的结构

α 区中的启动子 P 启始 S 和 U 或 S' 和 U' 基因的转录，gin 基因的产物作用于颠倒重复，使 G 区颠倒

根据对 Mu DNA 变性/复性过程的研究，其中 Mu 基因组 DNA 的单链能够互补和形成双链结构，而其中来自不同插入位点的染色体 DNA 则不能互补，仍保留单链状态，因此使这种 Mu DNA 形成分叉末端。电镜下通常只观察到分叉的右端，分叉的左端因太小而难以观察到。复性后的 Mu DNA 不仅含有一个分叉的末端，而且还含有一个约 3.0kb 的内部单链区，叫 G 区（图 2－29）。G 区呈环状，由噬菌体的一部分颠倒重复所形成。如果 G 区的末端发生缺失或与 G 区相邻的一段叫做 gin 的区域中发生点突变，从裂解的细胞中分

离的 Mu DNA 则不能形成 G 环。

G 区含有两套基因，即 Sv 和 U 或 Sv' 和 U'，后两个基因的转录方向与前两个基因的转录方向相反（图 2－30）。这 4 个基因都编码组成尾丝的部件。G 区可以进行位点特异性颠倒，颠倒后 U' 和 Sv' 基因转录，因为颠倒过程使这两个基因与其相邻的启动子位于同一条 DNA 单链上，另外两个基因则不能转录，因为它们的转录方向与启动子的作用方向相反。发生颠倒的区域位于两个长 34bp 的颠倒重复之间，颠倒过程主要由噬菌体基因 Gin 基因的编码产物即颠倒酶（invertase）催化，但还需要一种寄主细胞基因的产物。通过 DNA 片段颠倒来控制基因表达的现象还见于沙门氏菌（*Salmonella typhiimurium*）的 P1 和 P 7 噬菌体。

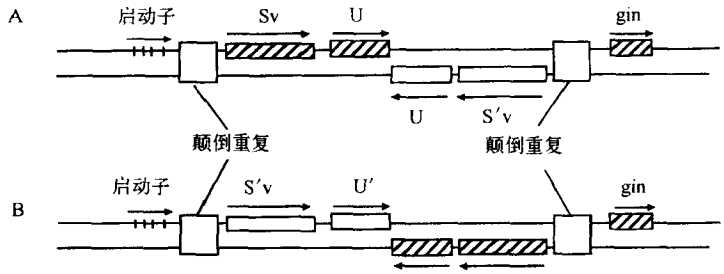


图 2－30 Mu 噬菌体的位点特异性颠倒

DNA 片段颠倒后，S' v 和 U' 基因转录 A. 顺方向 B. 颠倒方向

三、植物转座子

在高等植物中，研究最早和最为深入的转座子为玉米（*Zea mays*）的控制因子（controling element）。控制因子可以转座，影响基因表达，也能够引起染色体重组，包括缺失、重复、倒位和易位等结构变异。控制因子插入到某个基因座位上以后，在该座位上产生一种不稳定的等位基因。插入了控制因子的等位基因在体细胞和种质细胞中都不稳定，即隐性等位基因可以发生显性回复突变。

在植物中，各类控制因子都包括两个成员，其中之一为自主因子（autonomous element），即这种控制因子插入到某个座位上后，能够自动切割和再次转座，另一种为非自主因子（nonautonomous element），非自主因子只有在某种自主因子的帮助下才能切割和转座。自主因子为非自主因子提供切割和转座功能以及其他遗传重组特性。自主因子只能激活与自己属同一类型的非自主因子，而属不同类型的自主因子和非自主因子则不能互作。例如玉米中 Ac－

Ds 和 Spm－dSpm 就属于两类不同的自主因子和非自主因子系统。

（一）控制因子类转座子

玉米中激活子－解离（Ac－Ds）控制因子系统是研究最早和最为详细的植物转座子之一。解离（dessociation）是描述非自主因子的一种遗传学术语，因为这种非自主因子在属于同一类型的自主因子的帮助下，为染色体提供特殊的断裂位置或使染色体断裂成两条片段。激活子（activator）是和解离因子属同一类型的控制因子，能促进染色体在 Ds 位置断裂并使 Ds 转座，所以存在 Ac 因子时，由 Ds 引起的插入突变是不稳定的。此外，Ac 因子还可以独立地插入到某个基因座位上，产生不稳定的突变等位基因。

Ac 因子为一种结构和功能都比较完整的转座子，除了含有自动转座的信息外，还具有一些与切割和转座有关的基因或 DNA 序列，所以 Ac 因子的结构比较复杂，分子量也比较大。根据对玉米非蜡质基因（Wx）座位中 Ac 因子的序列分析，Ac 因子长 4 563bp，两端具有长 11bp 的颠倒重复序列，具有 2－3 个可读框，其中最大的一个（ORFa）可能有编码与转座有关的转座酶（transposase）（图 2－31）。在每一可读框之前存在 TATA 盖合，其后接有多聚 A 加尾信号序列。在 Ac 因子的插入位点上，两段颠倒重复序列之外各接有一段 8bp 的染色体 DNA 顺式重复。

Ds 为一种结构和功能都不很完整的转座子，只具有与切割有关的识别序列，缺乏与转座有关的功能，所以不能自动转座，容易在基因组中固定下来。根据各种 Ds 与 Ac 因子的序列同源程度差异，可以把 Ds 分成三种不同类型，第一类为 Ac 因子的缺失体，第二类 Ds 只在总长不足 1kb 的末端区同 Ac 因子存在同源性，而中部序列则与 Ac 因子无关（如 Ds2）。第三类 Ds 只是在末端颠倒重复序列以及 3’端约 20bp 的序列同 Ac 因子存在同源性（如 Ds1）。同 Ac 因子一样，Ds 也具有 11bp 的颠倒重复和 8bp 的染色体 DNA 顺式重复序列。

植物中另一类控制因子为抑制子－突变子（suppressor－mutator）系统，简称 Spm－dSpm 系统。这类控制因子也由两个成员组成，即一个长 8.3kb 的 Spm 因子和一个较小的、叫做 dSpm 的成员。dSpm 为 Spm 因子内部发生缺失后形成的转座子，但能够独立地转座。所谓抑制子（suppressor）即 Spm 插入到某个基因座位中后，往往抑制该基因表达，而突变子（mutator）则是指 Spm 转座所需的功能，它可能是指转座酶。Spm 因子具有上述两种功能，而 dSpm 则不具有抑制基因表达的功能，只有植物基因组中存在 Spm 因子时，dSpm 才能抑制插入位点基因的表达，即 Spm 以反式互补的方式为 dSpm 提供抑制功能。

各种 Spm 和 dSpm 因子的末端都具有 13bp 的颠倒重复序列，转座以后，都能在插入位点上产生 3bp 的染色体顺式重复。同其他植物转座子一样，Spm 转座以后不能使插入位点的基因完全恢复到原来状态，而是在该位点造成一小段重复序列。Spm 的主要转录子只有 2.5kb 长，编码一种反作用抑制子（trans－acting suppressor）蛋白，这种蛋白质也可能行使与 Spm 转录有关的一种正调节子的功能。Spm 也编码一些小转录子，这些小转录子含有主转录子的第一个内含子中的两个可读框即 ORF1 和 ORF2（图 2－31）。所有这些转录子都可能是前体 mRNA 通过不同加工途径形成的。当这两个可读框中发生缺失突变或移码突变之后，Spm 往往失去转座功能，而只具有抑制功能，因此，可能二者的编码产物都是转座酶的组成部分。

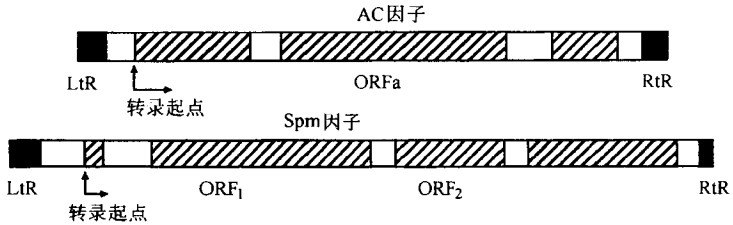


图 2－31　转座子 Ac 和 Spm 的结构模型

涂阴影区为可读框（ORF），Ltr 和 Rtr 分别为左、右两个颠倒重复

（二）非控制因子类转座子

在高等植物中，有些转座子只有一个成员，如金鱼草（*Antirrhinum majus*）基因组中的 Tam 转座子和玉米中的 Mu 转座子等。金鱼草中的 Tam1 转座子与 Spm 因子在结构上有些相似，二者的颠倒重复序列相同；转座以后都能在插入位点上产生 3bp 的染色体顺式重复 DNA，并且 Tam1 外显子的大小及组织也与 Spm 的相应结构相似。Tam1 中有一个可读框与 Spm 的 ORF1 具有 56％的序列同源性，可能这个可读框也编码转座酶的一部分肽链。而 Tam3 转座子则与 Ac 因子的结构相似，二者在两端有 7bp 的序列同源，转座后也都在插入位点上产生 8bp 的染色体 DNA 顺式重复，Tam3 唯一的一个无内含子的编码区也与 Ac 因子存在 50％～65％的氨基酸序列同源性。

四、酵母的转座子

酵母中最早发现的转座子为 Ty1 因子（Cameron，1979），此后又至少有三个转座子家族 γ、δ 和 σ 相继被发现。这些转座子在酵母基因组中的移动主要有两种方式，一种是重组，另一种是转座。

酵母的 Ty 因子长 6.0kb，包括 2 段长 334bp 的，叫做长末端重复（LTR）的顺式重复序列（图 2－32）。LTR 又叫做 δ 因子。δ 因子大约由 70% 的 AT 组成，每一个都含有一个启动子和一段被转座酶识别的序列。

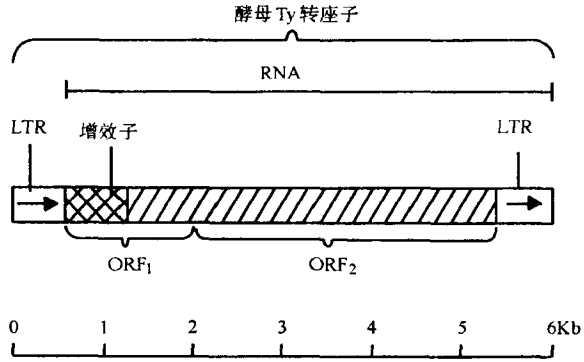


图 2－32　酵母 Ty 转座子的结构

根据对酵母组氨酸合成酶基因 his4 的研究，当 Ty1 插入到这一基因的启动子中之后，该基因不能转录，并在 Ty1 右端（下游）产生 5bp 的重复 TAAGA 序列。当 Ty1 从这一座位转移到新的座位以后，his4 基因便恢复正常表达功能。因此在这一点上，Ty 因子的行为特征和结构特征都类似细菌中的转座子。

由于每一 Ty 因子每一端都有一个 δ 因子，所以这两个 δ 之间就会经常发生同源重组，其结果就会产生一个游离的、具有一个 δ 的环状 Ty 因子并在染色体中留下一个 δ 因子。游离的 Ty 因子的命运尚不清楚。据研究，每一酵母细胞中有数百个 δ 因子。这些 δ 不能转座，但由于其含有转录启始信号，它们可能会影响其附近基因的表达。另外由于每一细胞中有多个拷贝的 Ty 因子，所以基因组中不同位置的 Ty 因子也可能发生重组，从而导致染色体的易位、缺失和倒位等结构变异。

Ty 因子只编码一条长 5700nt 的 mRNA，转录的启动子位于 Ty 因子左端的 LTR 之中。转录子含有两个可读框，所以这条 mRNA 可能编码两种蛋白质。在某些酵母品系中，Ty 因子可多达 35 个拷贝，但拷贝数因品系不同而有差别。

Ty 转座子中基因的转录比较复杂，根据对 Ty1 的研究，所有基因的转录都高度受到交配位点（mat）的影响（第八章）。在单倍体细胞或 a 和 α 纯合体中，细胞中大约有 5%～10% 多聚 A 化 mRNA 由 Ty1 基因编码，而在杂合体中，Ty1 转录子的含量则降低 75%～95%。交配型杂合体中 Ty1 基因表达水

平降低与酵母基因组中 SPT3 基因表达有关，如果这一基因发生突变就会抑制转座子的活性。转座子转录水平降低是由于位于 δ 单位左端的转录起点以下 800bp 的序列发生变化的结果。另外，培养基中的营养成分如 C 源也影响 Ty1 中基因的表达。例如甘油具有抑制 Adh2 基因（一种酒精脱氢酶基因）表达的作用，当 Ty1 插入到 Adh2 基因 5' 区域中后，甘油就成为 Ty1 表达的抑制物。相反，葡萄糖是 Adh2 基因表达的抑制物，但它在这里则成为一种脱抑制的活化物，使 Ty1 的表达恢复到正常状态。

Ty 因子转座是通过一种 RNA 中间产物进行的，其过程类似逆转录病毒（retrovirus）的复制与整合（第六章），所以 Ty 因子也叫做逆转座子（retrotransposon）。一般认为 Ty 因子转座时，首先以其 DNA 为模板合成一个拷贝的 RNA，然后再通过逆转录，合成一条新的 Ty 因子，最后这条新的 Ty 转座子再插入到新的染色体位点上。

五、无脊椎动物的转座子

有关无脊椎动物基因组中的转座子下面将以果蝇中的 Copia 因子和 P 因子家族为例作说明。

（一）果蝇的 Copia 转座子

在果蝇基因组中，Capia 为一个转座子家族，其成员的长度从 5 000bp 到 9 000bp不等，各个成员的两端都接有长 200～500bp 的顺式重复（LTR），这一特征与酵母的 Ty 转座子相同。每一成员两端顺式重复序列并非高度同源，其中存在一部分碱基不能互补的区域。在果蝇基因组中，Copia 约有 30 个拷贝，分布在几条不同染色体上，各个成员之间存在高度结构保守性。当 Copia 插入到某个基因座位上以后，也在其两侧形成 5bp 的染色体顺式重复 DNA。在成长的果蝇的细胞中以及在胚胎细胞中还存在多个拷贝的、呈小环状的、游离于染色体外的 Copia 转座子。

同酵母的 Ty 转座子相似，Copia 也是通过一种 RNA 中间产物进行转座。

在果蝇基因组中，还有一些转座子与 Copia 相似，它们都属于 Copia 转座子家族，如 B104、412 和 297 等。B104 转座子的末端顺式重复 LTR 与 Copia 的相应结构有一定程度的同源性，在转座过程中也产生源于染色体 DNA 的、长 5bp 的顺式重复。B104 长 8，700bp，在单倍体基因组中约有 80～85 个拷贝。转座子 412 两端也都有一段很长的 LTR（481bp），但在转座过程中只产生 4bp 的染色体顺式重复。也同 Copia 一样，在培养的果蝇细胞中，有一部分为游离的环状 DNA 分子。转座子 297 与 Copia 的同源程度不及上述两种转座

子高，其两端也各有一段长的 LTR，根据其三个不同插入位点的结构分析，转座子 297 都产生 6bp 的、相同的染色体顺式重复 TATATA 序列，可能这种转座子具有位点特异性。

（二）果蝇杂种不育与 P 转座子

杂种不育（hybrid dysgenesis）是普通果蝇的一系列缺陷的综合表现，包括突变、染色体畸变和不育等。例如将实验室培养的雌果蝇与果蝇的自然群体交配时就表现出杂种不育现象，这种实验室品系就叫做 M 细胞型（maternal contributing cell type），自然品系就叫做 P 细胞型（paternal contributing cell type），所以这种杂交组合就是 M♀×P♂。如果反交，P♀×M♂，所有后代都是可育的，不存在杂种不育现象。

杂种不育现象基本只影响生殖细胞，（M♀×P♂）F_1 杂种的体细胞正常，而性腺（gonad）不发育，所以成蝇是不育的。当 P 雄性亲本的染色体处于源于 M 雌性亲本的细胞质中时就会造成杂种不育。P 雌蝇的细胞质不能诱导杂种不育，所以 P♀×M♂的子代都完全可育。

诱发 M♀×P♂的杂种不育的模型如下：P 雌体的染色体载有 30～50 个拷贝的叫做 P 因子的转座子，而 M 品系则不含任何 P 因子。P 因子长 0.5～2.9kb，每一因子都有长 31bp 的末端颠倒重复。有些较短的 P 因子为非自主因子，这同玉米的 DS 因子相似，它可能源于较长的 P 因子。较长的 P 因子同玉米的 AC 因子相似，为自主因子，它编码一种转座酶，该酶催化自主因子的转座和较短 P 因子的转座。M♀×P♂的 F_1 代杂种的染色体含有插入到新位点上的 P 因子，其原因是 P 因子的转座过程被活化，而这种活化过程只发生在生殖细胞中。

P 转座子的转座具有组织特异性，它只在种质细胞中转座，但是它在体细胞和种质细胞中都能进行转录。这种组织特异性是由 RNA 的可变剪接（differential splicing）控制的。P 转座子只有一个编码区，由 4 个外显子和 3 个内含子组成。初级转录子长 2.5～3.0kb，可编码两种蛋白质（图 2－33）。在体组织中只除掉开头的两个内含子，形成由 E1、E2 和 E3 组成的编码区。这种 mRNA 通过翻译，合成 66kD 的抑制转座的抑制蛋白。在种质细胞中，剪接过程将 3 个内含子都除掉，形成由 4 个外显子组成的 mRNA。这种 mRNA 编码出 87kD 的蛋白质即转座酶。

P 转座子转座的关键是第三个内含子是否被剪接掉。如果这个内含子被除掉，4 个外显子就都能连接在一起，P 因子就能转座。体细胞中有一种蛋白质调节 P 转座子的组织特异性剪接，这种蛋白质与第 4 个外显子结合，阻止最后一个内含子 I3 的剪除。由于种质细胞缺少这种蛋白质，所以其中的初级转

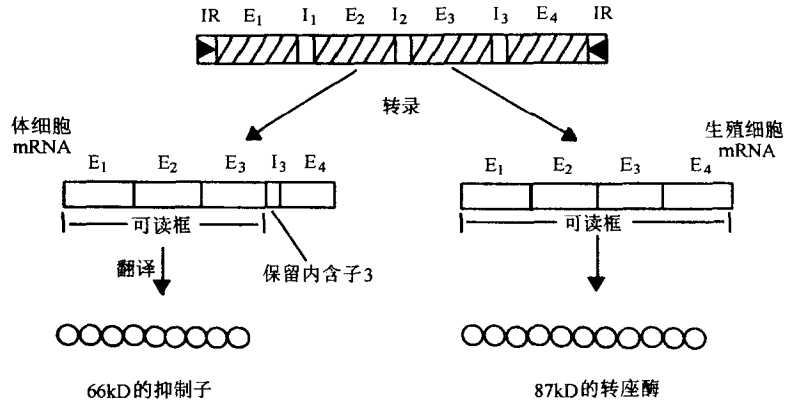


图 2－33 果蝇 P 转座子的结构及其编码产物

P 因子有 4 个外显子，在体细胞中前 3 个外显子被剪接在一起，在生殖细胞中，所有 4 个外显子被剪接在一起

录子通过剪接就能形成编码转座酶的 mRNA。

P 转座子转座需要其末端的一段 150bp 的序列。转座酶与其中 10bp 的、与 31bp 的颠倒重复相邻的一段序列结合。其转座过程属于非复制型转座，当转座子从供体位点上切割下来后，通常在供体位点上形成缺口，造成染色体断裂。在某些情况下，供体 DNA 的断裂可利用同源染色体的序列为模板进行重组修复。但是如果同源染色体含有一个 P 转座子，通过修复，供体位点又可重新获得一个 P 转座子，这一过程与复制型转座相似。如果同源染色体不含转座子，那么供体位点就可得到较完整的修复。

进一步研究知道，果蝇的杂种不育既依赖于 P 转座子本身，同时也依赖于杂交亲本的细胞质。这种细胞质效应与 66kD 的抑制蛋白有关。抑制蛋白既能抑制转座酶合成，又能抑制转座酶的活性，它贮存在卵细胞质中，表现为母体效应。所以无论在 P♀×P♂杂交中还是在 P♀×M♂杂交中，由于卵细胞质中都存在抑制蛋白，P 转座子不能转座，杂种都可育。只有在 M♀×P♂杂交中，由于卵细胞质源于 M 细胞型，缺少抑制蛋白，当转座酶合成后，就活化 P 因子转座，从而导致杂种不育现象。

P 因子可以用作载体，将外源基因转移到果蝇胚胎的生殖系中，对果蝇进行遗传操作（见第九章）。

果蝇基因组中除了存在 Copia 和 P 转座子外，还存在其他几种转座子家族，如 R 家族、hobo、opa 和 mdg 转座子等，这里不作详细介绍。

六、转座子转座与基因突变

（一）细菌和植物转座子的转座模型

关于转座子转座的机制目前有多种假说，因为对不同生物乃至同一生物中不同类型的转座子来说，可能转座机制也都不一样，因此不可能用一个统一模型加以说明。这里只对细菌和植物中有关转座子转座的几种机制加以讨论。

1. *细菌转座子的复制转座*　有关细菌转座子的复制转座过程目前主要的有两种模型，即首先复制模型（replication－first model）和首先重组模型（recombination－first model）。这两种模型主要是依据转座子如下几个共同特征提出的，（1）各种细菌转座子的末端都存在颠倒重复；（2）转座之后，都在插入位点上产生染色体顺式重复；（3）转座期间都形成一个拷贝的转座子；（4）都能形成共整合体（cointegrate），并造成缺失（deletion）和倒位（inversion）。一般认为，转座子的颠倒重复是转座酶的识别、切割和使转座子连接到插入位点上的功能结构；插入位点上的顺式重复则是转座酶交错切割后，DNA 修复合成所形成的；而缺失和倒位则是分子内重组的结果。

首先复制模型认为，转座子在切割和连接到目标位点上之前进行复制，这种模型起初是根据 Mu 噬菌体的切割和整合过程提出的，但是目前还没有足够证据支持这种假说。首先重组模型认为，插入位点上的交错切口与转座子两端连接时，造成整个转座子序列发生重复。首先重组模型又分为对称模型和非对称模型，前者认为转座子两端同时分别切割两条单链并连接到目标位点的 DNA 分子上，而非对称模型则认为转座过程只由转座子的一端与目标位点连接而引起。因此，在对称模型中，共整合体形成是一个必需的中间步骤；而在非对称模型中，共整合体则是一种附产品（图 2－34）。根据对 Tn3 的研究，对称模型中共整合的形成由基因 tnpA 的编码产物所催化，而其溶解则由 tnpR 基因产物所控制。

上述两类复制转座模型中，转座子只将其一个拷贝转座到新位点上。在原核生物中还同时存在另一种不同的转座机制，即转座子从原来位置上切割下来，然后再插入到新的位点上（如 Tn5 等），因此这种转座机制不可能提高基因组中转座子的拷贝数。

2. *植物转座子的非复制型转座模型*　在植物中，当转座子插入到某个基因座位以后，往往造成不稳定的隐性突变，也就是说隐性突变等位基因可以恢

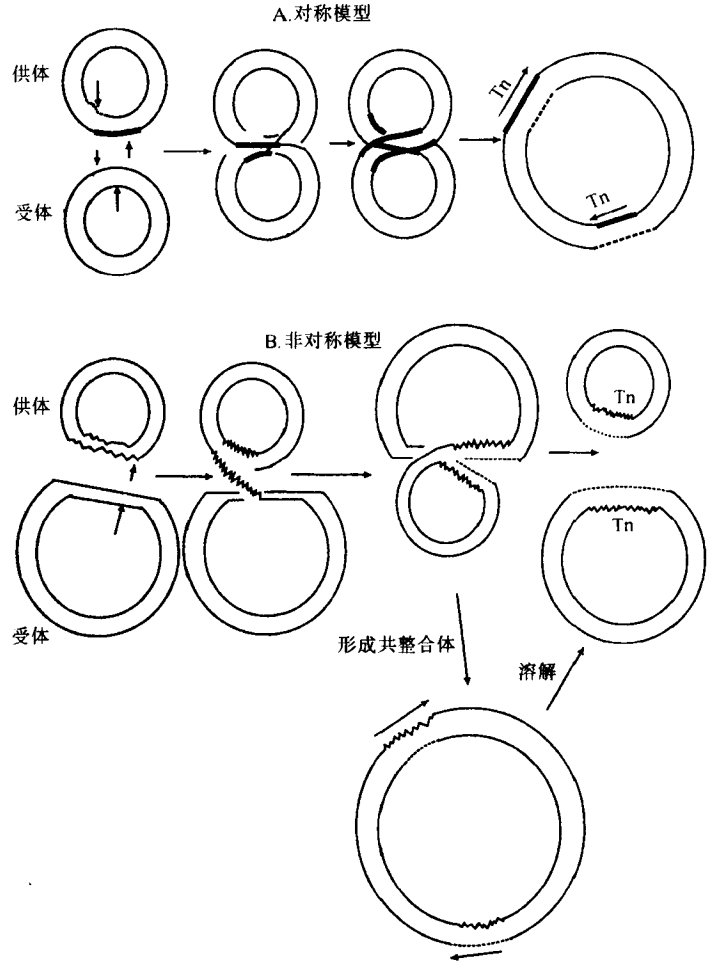


图 2－34　细菌转座子转座的对称和非对称模型

在对称模型中（A），转座子两端与目标位点的 DNA 序列连接，转座子发生重复，形成中间产物共整合体，当共整合体溶解后，转座过程完成；在非对称模型中（B）转座子通过滚环复制模型插入到目标位点上；共整合体可以通过两轮 DNA 复制而形成，也可以通过单链交换形成

复到显性状态，表明植物转座子首先从基因组中某个位置切割下来，然后再插入到另一位点上。以玉米 Spm 转座子为例，这种转座子编码两种功能，即抑制子和转座酶。一般认为在转座过程中，转座酶识别转座子的两个末端颠倒重复并将其结合在一起，使二者互作；然后转座酶启始转座子两侧的染色体顺式重复的切割过程，使之形成交错末端（图 2－35）。这一过程可能涉及几种酶，

这些酶与DNA结合，形成半稳定的蛋白质/DNA复合体以防止转座子在切割后立即释放下来。然后，蛋白质－DNA复合体中3′端和5′端由修复酶进行修复；在这一过程中，一种5′外切核酸酶可能降解其侧翼染色体DNA的5′单链末端序列。转座子的5′端可能从复合体上暴露出来，因此也易于受到外切核酸酶的作用。降解作用可能在双链DNA位置终止，同时DNA聚合酶也可能与染色体DNA的游离3′端结合，利用顺式重复的单链末端为模板修复缺口。

在这种模型中，如果一条顺式重复一端的单链区完全降解，而另一重复的一端则不降解，其结果就使染色体恢复到原来的结构。如果一条顺式重复的一条单链末端完全降解，另一重复的单链末端也完全降解或部分降解，这样就会造成染色体缺失。

根据切割过程中修复酶作用的准确性和修复程度以及外切核酸酶降解单链末端的程度，切割后就产生各种结构变异（图2－35）。

转座子从基因组中某个位点切割下来后，如果其5′端的单链序列完全被外切核酸酶降解，转座子将不含任何外源DNA序列。如果末端的单链区只是部分被降解，那么转座子的一端或两端仍接有数目不同的碱基。转座子上的这类外源DNA序列使转座子在下一轮转座过程中具有插入位点特异性，即转座子趋向插入到与这类侧翼序列存在同源性的某个位点上。

一般认为，转座子的插入过程也由转座酶所催化，与转座子末端具有亲和性的转座酶也同样交错切割目标插入位点；如果转座子不含有原来插入位点的额外核苷酸，那么转座子的3′端直接与插入位点序列的5′端连接，转座子5′端与插入位点序列3′端之间的缺口则

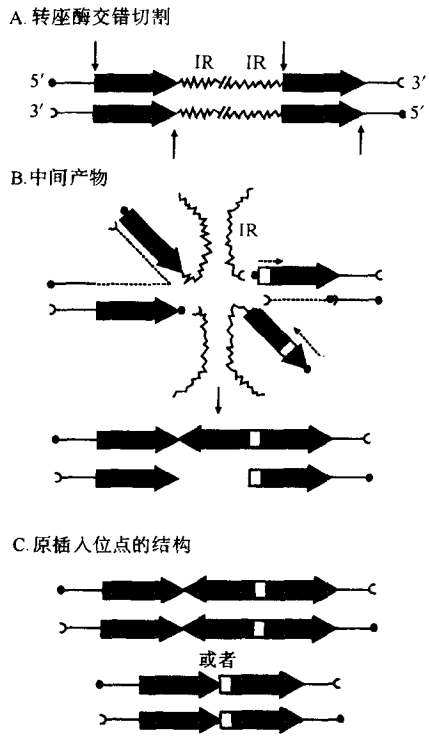


图2－35　植物转座子的切割过程

波浪线示转座子，其末端由颠倒重复序列（IR）组成；直线和黑箭头分别示转座子插入位点的染色体DNA和顺式重复，重箭头中未涂黑的区域为缺失序列；细垂直箭头示转座酶的切割位点；B中细虚线箭头示外切核酸酶降解区，粗虚线示DNA聚合酶的修复区

由DNA聚合酶通过修复合成来填补。因此，转座子转座后只是在新的插入位点产生顺式染色体重复(图 2－36B)。在这个过程中，5′外切核酸酶也可能降解插入位点的一部分单链序列，使转座子两侧的染色体顺式重复不完全相同。

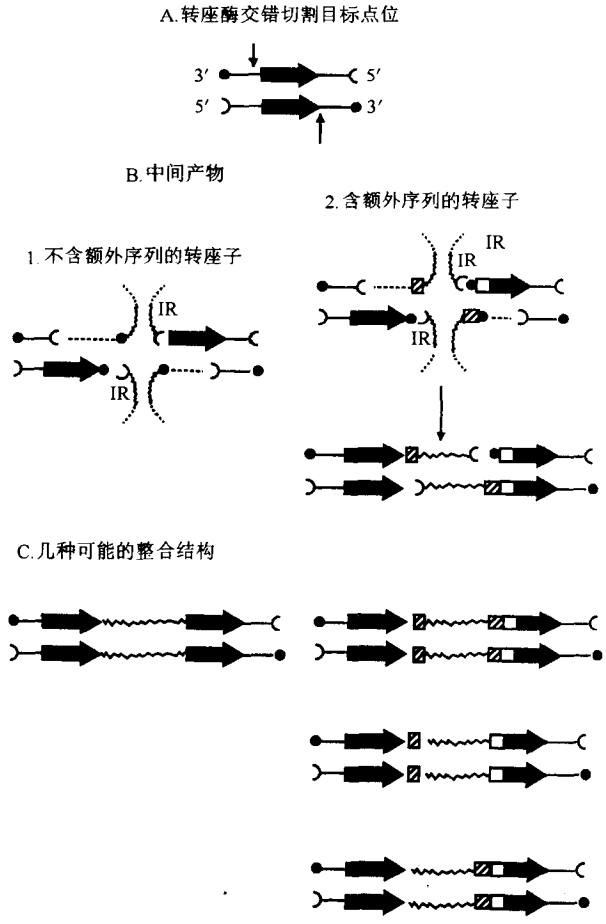


图 2－36 植物转座子的插入过程

A. 粗水平箭头示染色体目标位点的序列，转座酶在这一位点交错切割（细垂直箭头）B. 整合过程中两种可能的中间产物，粗虚线示从5′端开始的修复合成，转座子两侧粗箭头中未涂黑部分示由外切核酸酶降解的缺失部分 C. 修复合成或DNA复制之后，可能形成的各种产物

如果转座子的5′端载有额外核苷酸，当3′端与插入点的5′端连接后，这

些额外核苷酸就可能形成不配对的单链环，这种环状结构既可以通过修复合成而被除掉，也可以通过合成相应的互补链而整合到基因组中。另一方面，由于转座子5′端带有额外核苷酸，这些额外核苷酸可能会阻止其3′端与插入位点的5′端迅速连接，使插入位点的5′端受到外切核酸酶的降解，从而导致一条顺式重复中发生缺失（图2－36）。

当然，这种转座机制并不能代表植物中所有转座子都按这种模式转座。在这个模型中，转座子只是从基因组的一个位置转移到另一位置，转座过程并不进行DNA复制，可能不同植物的不同转座子的转座方式各不相同，有些可能只是在DNA复制过程中进行转座，也只是将其一个拷贝转移到新的座位上如Ac因子，从而使基因组中转座子的拷贝数增加（图2－37）。

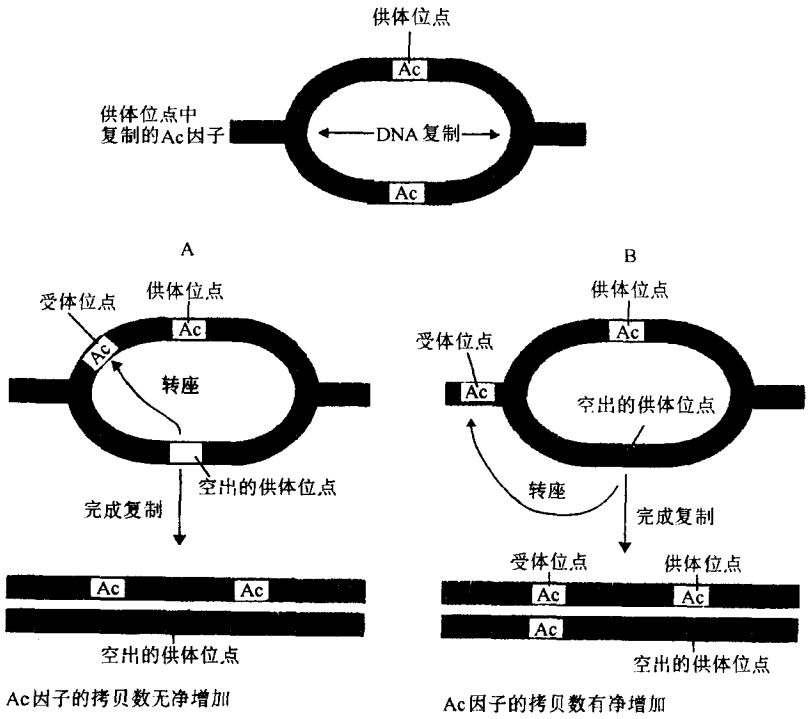


图2－37　Ac因子的转座机制

A. 转座到已复制的受体位点上，基因组中Ac因子的拷贝数无净增加　B. 转座到未复制的受体位点上，当含有转座子的染色体区复制后，基因组中Ac因子的拷贝数就有净增加

（二）转座子转座与基因突变

已如本节前述，当转座子插入到某个基因座位中后，往往严重影响该基因

表达，如植物转座子插入到某个基因中后，常产生不稳定突变等位基因，当从该座位上切割下来后，该基因又可恢复到显性状态。但是，根据转座子在基因中插入位点不同，影响基因表达的效应也有很大差别。

根据对许多植物转座子的研究，当转座子插入到某个基因的外显子或非翻译前导序列中以后，往往导至该基因失活，但在某些情况下，插入位点的基因仍然能够正常转录，只是转录子中的插入序列通过转录后的剪接过程而被除掉，因此插入位点的基因仍表现出显性性状，这种现象叫做泄漏突变（leaking mutation）。

插入失活和泄漏突变与转座子在插入位点中的方向有关。例如，当 dSpm 和 Ds 因子与所在基因的方向相反时，前体 mRNA 中的 dSpm 或 Ds 序列则可通过转录后的加工过程而被除掉，所以含有转座子的基因仍然编码野生型蛋白质和 mRNA，即所谓的泄漏突变。如果转座子在外显子中的转录方向与所在基因的转录方向相同，转录过程常终止在转 座子的多聚 A 化信号位点上，形成半截子 mRNA，因此也造成插入失活。

对于 Spm/dSpm 控制因子系统来说，由 dSpm 造成的泄漏突变只是在基因组中不存在 Spm 因子时才能发生，如果植物基因组中存在 Spm 因子，无论 dSpm 的插入方向如何，含有转座子的基因都不能正常表达。这种现象进一步说明，Spm 以反式互补方式为 dSpm 提供抑制功能。

在细菌中，根据对乳糖操纵元和半乳糖操纵元的研究，当 IS 因子插入到这两种操纵元的 5′端顺反子中以后，往往严重降低其下游顺反子的表达水平，这就是所谓的极性突变（polar mutation）或突变的极性效应（polar effect）。

在真核生物基因组中，当转座子插入到某个基因的内含子中以后，也会影响该基因表达，通常表现为降低基因的转录水平，但在某些情况下也会改变内含子的剪接位点，使某些外显子丢失，因此也导至基因失活。

如果转座子插入到某个基因的转录控制区或启动子中，则导致基因完全失活；当转座子切割以后，由于往往造成插入位点发生缺失或重复等结构变异，因此转录控制区在结构和功能方面就不可能完全恢复到原来的状态，而使基因的转录发生各种不同程度的变化，或者严重降低基因的转录水平，或者改变基因的组织、器官特异性表达等。

插入到某个基因座位中的转座子也有可能失活，失活过程主要与 DNA 分子的甲基化作用（methylation）有关。当转座子中的某些甲基化位点脱甲基化以后，转座子又可以恢复转座功能，使转座子表现出所谓的外遗传变异（epigenetic change），即某个等位基因的表现型发生改变，而其基因型保持不变的

现象。转座子 DNA 序列的甲基化不仅影响转座子本身，而且还影响邻近基因的表达，使邻近基因也表现出外遗传变化。

小　　结

基因是 DNA 分子中一段序列，它编码某种蛋白质，某种多肽或某种直接参与细胞内代谢活动的 RNA。RNA 聚合酶可以将其中一段序列转录成与之互补的 RNA，这段序列叫做转录单位。所有基因都有控制基因表达的一些特殊序列，这类序列统称为控制序列，包括启动子、终止子、核糖体结合位点、增效子等。编码蛋白质的基因和 rRNA 基因的启动子位于转录起点的 5′端。原核生物基因的启动子有两段结构保守序列，一个为 - 10 位的普里布诺框 TATAAT，另一段为 - 35 位的 TTGACA 序列。真核生物基因的启动子是指准确而有效地启始基因转录所需的任何 DNA 序列，其中保守性最高的为 TATA 框。终止子是终止转录的信号，核糖体结合位点是转录单位内、mRNA5′端序列中与核糖体小亚基结合的一段序列，真核生物蛋白质基因还有控制 mRNA3′端加尾的信号。在某些小 RNA 分子和某些病毒基因中，启动子和转录终止信号位于转录单位内。

真核生物的大多数蛋白质基因都是二型蛋白基因，含有两个连续的编码区，编码两种不同功能的多肽。有些二型蛋白基因编码的功能部分隐蔽在前体蛋白之中，在细胞需要时才经过酶切、加工，释放出功能蛋白。还有某些基因缺少完整的编码序列，但这些基因通过转录后的 RNA 编辑，使 DNA 分子中不完整的遗传信息在 RNA 水平上保持完整。在脊椎动物的基因组中，与免疫有关的基因不能直接转录，须通过一系列重组过程，将各条 DNA 片段组装成能够转录的功能基因。

原核生物的大多数基因以及真核生物的 rRNA 基因都以操纵元的形式组织在一起，受一个转录开关控制，转录出一条很大的多顺反子 RNA，然后多顺反子 RNA 或者通过翻译，产生几种不同的多肽，或者通过切割、加工，形成几种不同的 RNA。

各种生物的基因组中都有一些基因为单拷贝基因，没有重复序列，但还有一些基因有两个以上拷贝即重复基因，它们或者首尾连接在一起，形成基因簇，或者分散在整个基因组中。此外，基因组中也存在重叠基因，即两个以上基因共用一段 DNA 序列作为转录的模板。

基因可以发生突变。突变的机制可以分成两大类，第一类是碱基的转换和颠换，第二类是核苷酸的插入和缺失。转换和颠换会造成误义突变、无义突

变，使基因的特性改变；但也会造成中性突变或同义突变，并不影响基因的特性。单个碱基的插入和缺失可使基因的编码序列移码，使原来编码某种多肽的序列改变成编码另一种多肽的序列。细胞中有一些移码抑制因子即一种特殊的tRNA，它们可以识别由4个以上碱基组成的密码，抑制移码突变。

基因突变可以通过某些化学因素和物理因素进行人工诱变。各种化学诱变剂有三大类，即碱基类似物、碱基修饰剂和嵌入剂。由碱基类似物和某些碱基修饰剂诱发的突变可以第二次用同类诱变剂处理，诱发回复突变。利用DNA重组技术，可以在某个特殊基因的特定核苷酸位置创造位点特异性突变，然后再将突变基因送回细胞中，研究突变的效应。

基因重组是两条DNA分子发生片段互换的结果。重组的断裂重接模型认为，减数分裂过程中只有具有同源性的染色体之间才能发生断裂重接。霍利迪模型认为，重组是由两条双链DNA分子中发生单链断裂所启始的。这两个断头分别侵入相对应的双螺旋亲链中，形成一段杂双螺旋。两条双螺旋亲链通过霍利迪连结体连接在一起，当切割霍利迪连接体以后，通过修复便形成两条互换了片段的完整的双螺旋。重组期间DNA杂双螺旋通过复制或通过误配对修复转变成纯双螺旋，从而使一个基因获得另一基因的序列，最后造成基因转换。

有关参与重组的酶类，已知的主要为大肠杆菌中由rec座位和ruv座位编码的蛋白质。其中RecA蛋白能使一条双螺旋的一条单链侵入到另一条双链DNA中，将两条同源DNA分子联会在一起。RecBCD核酸酶结合到DNA的chi位点上将DNA解螺旋并切割chi序列中一条单链，所以chi位点是重组的关键。RuvA和RuvB蛋白作用于杂双螺旋，RuvC切割霍利迪连接体。

原核细胞和真核细胞含有一些可以从染色体的一个座位转移到另一座位的DNA序列，这些DNA序列统称为转座子或转座因子。转座子在两端一般都有颠倒重复序列，并且可以在插入位点上产生两段染色体顺式重复。最简单的转座子是细菌的IS因子，它基本上只由两段末端颠倒重复和一个编码转座酶的可读框组成。复合座子如Tn5则在其两端接有IS因子，其中一个或2个IS因子为转座过程提供转座酶，即两个IS因子夹住中间一段主序列转座。IS和Tn5都是非复制型转座，即转座子从供体位点转移到受体位点。Tn3转座子和Mu噬菌体为复制型转座，当供体位点上的转座子连接到受体位点上后，通过DNA复制于是产生一个含有2个拷贝的转座子的共整合体。共整合体通过2个转座子之间的重组被溶解成2个独立的复制子，每一复制子都含有一个拷贝的转座子。这类转座子编码两种酶，一种是转座酶，它识别转座子的两个末端并将这两个末端连接到靶位点上。另一种是溶解酶，它催化位点特异性重

组，溶解共整合体。

植物中研究较清楚的转座子为玉米的控制因子，包括几个不同家族，每一家族都有一种能够自动转座的自主因子和多种不同的非自主因子。非自主因子不能自动转座，只有存在同类自主因子时才能转座和表现出自主因子的其他特征。

植物如玉米的转座子的转座过程属于非复制型转座，可能只需要由转座子本身编码的一种转座酶。如果转座子随着所在染色体的区域复制以后再转座到未复制的染色体区域，当染色体复制完成后，就有可能增加基因组中转座子的拷贝数。非复制型转座子从原位转座到新的靶点上后，会使染色体发生缺失、重复、倒位、易位等结构变异。

普通果蝇的P转座子与杂种不育有关。载有P因子的雄蝇与缺少P因子的雌蝇杂交，其子代表现不育。P转座子只有一个可读框，有4个外显子。剪接过程中，如果开头的三个外显子连接在一起，编码66kD的蛋白，它存在于所有细胞中。如果4个外显子都连在一起，则编码87kD的转座酶。第二种剪接方式具有组织特异性，只发生在生殖细胞中。只有细胞质缺少抑制蛋白时，P因子才能转座。如果P因子转座，杂种表现不育。只有完整的P因子才能编码转座酶，而缺陷型的P因子可以通过转座酶的反式作用进行转座。

酵母的Ty因子和果蝇的Copia转座子在两端都接有长末端重复LTR，LTR中含有控制转录和转座的信号。这两种转座子转座时，都首先以DNA为模板合成一个拷贝的RNA，然后再以RNA为模板合成一条新的DNA，最后这条新的转座子再插入到新的染色体位点上。所以这两种转座子又叫做逆转座子。

复 习 题

选择1～3题中正确答案：

1. DNA单链断裂或双链断裂都可能诱发重组

A. 一条单链断裂是吸收另一条同源单链所必需的。

B. 一条单链断裂是两条双螺旋DNA之间进行单链互换所必需的。

C. 双链断裂启始供体双螺旋的类似D环的单链替换和启始下一步单链DNA合成。

D. 为避免丢失遗传信息，双链断裂的修复需从另外的同源双螺旋中获得序列。

E. 双链断裂是启始由Rec BCD所催化的重组过程所必需的。

2. 复制型转座是

A. 转座子重复，在原位留下一个拷贝。

B. 转座子从一个位置转移到另一位置，原位没有转座子。

C. 需要转座酶。

D. 需要溶解酶。

3. Copia 转座子

A. 具有 LTR。

B. 以衔接重复方式组织在一起。

C. 两侧接有顺接重复。

D. 并不转录。

4. 比较前导序列、拖尾序列、编码序列，内含子、间隔序列、内部转录间隔序列的差别。

5. 真核生物蛋白质基因的启动子区中最基本的结构保守序列有哪两种？

6. 原核生物基因的 5′转录控制区由哪两种结构保守序列组成？

7. 隐密基因同隐形蛋白基因有什么本质区别？

8. 哺乳动物的免疫球蛋白基因表达时，能否直接以染色体 DNA 为模板转录？为什么？

9. 原核生物操纵元（操纵子）的转录控制区中有哪三种重要结构？它们分别与哪些蛋白质互作？

10. 如果三个不同基因转录时共用同一段 DNA 序列，它们称为

A. 基因簇　　B. 重复基因　　C. 重叠基因

11. 点突变和移码突变的遗传效应有什么不同？

12. 某种蛋白质的正常序列以及 4 种突变体的序列如下：

正常：Met－Gly－Glu－Thr－Lys－Val－Val……Pro

突变体 1：Met－Gly

突变体 2：Met－Gly－Glu－Asp

突变体 3：Met－Gly－Arg－Leu－Lys

突变体 4：Met－Arg－Glu－Thr－Lys－Val－Val……pro

指出每一种突变体是在基因的编码序列中发生何种突变的结果。

13. 大肠杆菌的突变体 lacZ－1 是用吖啶染料处理细菌细胞诱导的，而 lacZ－ 2 则是由 5BU 诱导的，这两种突变体可能属于哪种类型的突变？试说明之。

14. 培养后的大肠杆菌细胞经 5BU 处理后，发生突变的频率比正常条件下突变出现的频率高出许多倍。将突变的菌落再用亚硝酸处理，其中有些突变

回复到野生型。

（1）根据碱基配对模型，图示 5BU 诱导突变发生的一系列步骤。

（2）假定回复突变不是由突变抑制剂引起的，说明亚硝酸诱发回复突变的步骤。

15．一种确定突变剂作用方式的试验是利用另一突变剂来确定一种突变剂的诱导产物的回复突变性。下表是各种突变剂对 T2 噬菌体的 rⅡ突变的回复性的资料。“+”示有回复突变，“-”示无回复突变，BU 示 5 溴尿嘧啶，AP 示 2-氨基嘌呤，NA 示亚硝酸，HA 示羟氨。请对下表填空。

突变剂	回复突变				碱基替换形式
	BU	AP	NA	HA	
BU	+			-	
AP		-		+	
NA	+	+		+	
HA			+	-	

16．大肠杆菌的 3 个 ara 突变体是由突变剂 X 诱导的，对另一些突变剂诱导回复突变的能力（ara-→ara+）进行了测定，其结果如下表

处理后 ara+ 细胞出现的频率

突变体	突变剂				
	无	BU	AP	HA	移码
ara-1	1.5×10^{-8}	5×10^{-5}	1.3×10^{-4}	1.3×10^{-8}	1.6×10^{-8}
ara-2	2×10^{-7}	2×10^{-4}	6×10^{-5}	3×10^{-5}	1.6×10^{-7}
ara-3	6×10^{-7}	10^{-5}	9×10^{-6}	5×10^{-6}	6.5×10^{-7}

假定所有 ara+ 细胞都是回复突变体，问哪些碱基变异可能与形成三种原始突变有关？突变剂 x 诱发哪些种类的突变？

17．一个 IS 因子插入到大肠杆菌的 lacZ 基因之中，造成该基因发生一小段缺失，即从 IS 左端开始丢失 40 对碱基，包括 lacZ 的 10 对碱基和 IS 最左端的 30 对碱基。这种缺失将会造成什么结果。

18．RecA 蛋白、RecBCD 酶和 Ruv 酶在重组过程中的主要作用各是什么？

19．为什么说基因转变总是与重组相伴发生？二者之间有什么区别？

20．基因转变极性是指

A．几个紧密连锁的基因中上游基因发生转变的频率高于下游基因。

B．同一基因座位内，5′位点发生转变的频率高于 3′位点发生转变的频

率。

21. 当 DNA 发生双链断裂后，其遗传信息是如何得到恢复的？

22. 大肠杆菌的 Tn3 和 Tn5 的结构和转座机制有什么不同？

23. 什么叫杂种不育？造成杂种不育的原因是什么？

24. 玉米的 Ac 因子转座时只是从一个位点切割下来再插入到另一位点上，为什么基因组中转座子的拷贝数会增加？

25. 逆转座子在结构上和转座机制方面同其他转座子有什么不同？

26. 转座子插入到新的位点上后，对插入位点的基因及其周围的基因有何影响？

27. 转座子转座后会引起染色体发生哪些结构变异？

第三章

原核生物基因组及其表达

原核生物为单细胞生物，其细胞结构简单，不具有细胞核和细胞器如线粒体和叶绿体等，只有一个通过浮尔根反应（Feulgen reaction）着色的区域，叫做核区（nuclearregion）。就细菌而言，各种细菌都含有一条很大的环状双链DNA分子，叫做细菌染色体，其中载有细菌进行生命活动的所有遗传信息。除了细菌染色体以外，许多细菌还具有一些小环状DNA分子，叫做质粒(plasmid)。质粒能够独立复制，细菌丢失质粒后，其生长和繁殖并不受影响。虽然细菌细胞结构简单，但它能够进行基本的生命活动。

第一节　细菌基因组及其复制

一、细胞中的拟核

细菌细胞中的基因组DNA为一种很大的环状双链DNA分子，它可以收缩为原来的千分之一，形成叫做拟核（nucleoid）的结构。拟核在功能上与真核细胞的细胞核相似，但在结构上有明显区别。拟核不由膜包裹，其稳定性比真核细胞的染色质低得多。

在电镜下，拟核为分裂状结构，其边缘为单链DNA，含有新合成的RNA和核糖体，所以边缘区域是拟核代谢非常活跃的区域。拟核的内部结构目前还不清楚，可能DNA是运动的，也可能拟核裂隙很深，基因组中任何区域都暴露在外。

拟核中的DNA为负超螺旋，闭合环状DNA（closed circular DNA），具有较高的自由能。负超螺旋DNA在形成单链缺口后可变成松弛型DNA（图3-1），但对整个细菌染色体而言，某个位点形成单链缺口并不能使整个超螺旋完全松弛。细菌染色体可分成许多拓扑学上不同的部分，电镜下每一部分都是一个很大的环，只有每一部分都有一个单链缺口，染色体DNA才能完全松弛。在生活的大肠杆菌细胞中，每一基因组DNA分子约有50个不同部分，造成

各个部分拓扑学差异的原因目前不清楚。一般认为，当 DNA 在修复或复制过程中形成缺口时，DNA 的大部分区域都处于超螺旋的浓缩状态。

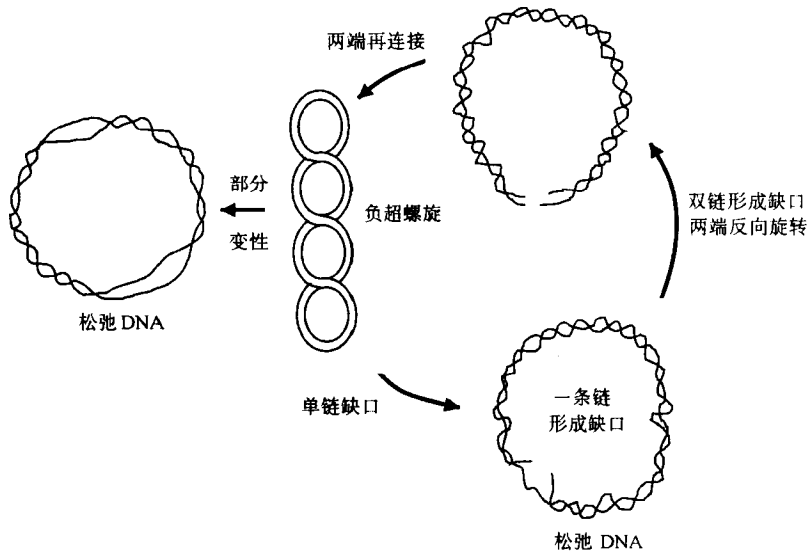


图 3－1　超螺旋和松弛 DNA 的形成

细菌 DNA 也与某些特殊蛋白结合，形成类似真核生物染色质的结构，但结合的稳定性很低。这些蛋白质类似组蛋白，其中最为清楚的为 Hu 蛋白。Hu 蛋白的分子量为 9 700，为一种热稳定性很高的碱性 DNA 结合蛋白，其物理特性和氨基酸组成都与真核细胞的组蛋白相似。Hu 蛋白在各种细菌中都是保守的，在离体条件下，Hu 蛋白能够将DNA 包裹成类似真核细胞核小体的颗

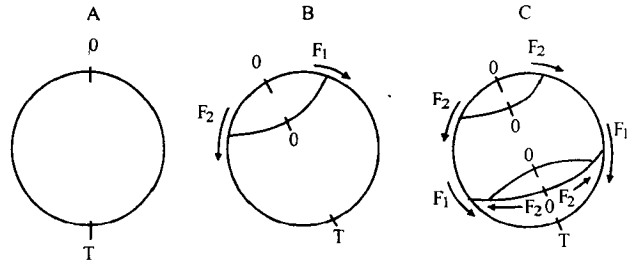


图 3－2　细菌染色体的复制

A. 复制首先从叫复制启始区（O）的位点开始，在叫复制终点（T）的位点终止

B. 两个复制叉（F）使染色体形成一种 θ 形结构，复制叉从复制起点朝复制终点双向前进

C. 当前一轮复制尚未完成时，第二轮复制就已开始，从而形成一种复合结构 F_1 和 F_2 分别为第一轮和第二轮复制的复制叉

粒。大肠杆菌中一些与 DNA 结合的蛋白质主要为 Hu 蛋白和另一种叫做蛋白质 I（分子量 16kD）的蛋白质。但目前还不清楚这些 DNA 结合蛋白质如何沿 DNA 分布。

细菌染色体上只有一个复制启始区（oriC），复制过程是朝复制终点双向进行的，所以一个细菌染色体就叫做一个复制子。在迅速生长的细胞中，当前一轮 DNA 复制尚未完成时就已开始下轮复制，因此这些细胞中的染色体往往是一种聚合体（图 3－2），形成多基因组细胞。

二、细菌 DNA 复制的主要酶类

细菌 DNA 的复制过程受许多酶或蛋白质因子的调节，但目前对它们控制染色体复制过程的确切机制还不很清楚。根据对大肠杆菌的遗传和生化研究，至少有 18 种不同基因的产物直接参与细菌染色体的复制过程（表 3－1），可能还有另外 14 种蛋白质（表 3－2）在细菌 DNA 复制过程中起着某种作用。

表 3－1　大肠杆菌中主要的 DNA 复制蛋白

蛋　白　质	基　本　功　能	分子量（kD）	亚基数	基　因
DnaA	识别复制起点	52.5	1	dnaA
DnaB	DNA 螺旋酶，合成引物	52.3	6	dnab
DnaC	将 DnaB 释放到 oriC	29	1	dnaC
SSB	与 ssDNA 结合	18.8	4	ssb
引物酶	合成引物 RNA	65.6	1	dnaG
DNA 聚合酶Ⅲ（全酶）	DNA 复制的聚合酶	740		
α	DNA 聚合酶	130	(2)	pdc（dnaD）
ε	3’－5’ 校对外切核酸酶	27.5	(2)	dnaQ（mutD）
θ	不清	10	1	未知
β	DNA 聚合酶附属蛋白	40.6	(2)	dnaN
τ	DNA 聚合酶附属蛋白	71.1	(2)	dnaZX
γ	DNA 聚合酶附属蛋白	52		dnaZX
σ	DNA 聚合酶附属蛋白	32		dnaZX
DNA 回旋酶（gyrase）	DNA 解螺旋，分离子染色体对	374	4	
α	闭切口	97	2	gyrA
β	ATP 酶	89.8	2	gyrB
DNA 聚合酶 I	消除引物，填补缺口	102	1	polA
DNA 连接酶	封闭 DNA 切口	74	1	lig
DnaT	终止 DNA 复制	19.3	3	dnaT
RNA 聚合酶	活化 oriC，转录	460	5	rpoAB
拓扑异构酶Ⅳ	分离子染色体			rpoC、D

表 3-2　大肠杆菌染色体复制有关的辅助酶

蛋　白　质	已　知　功　能	分子量（kD）	亚基数	基　因
Rep 蛋白	DNA 螺旋酶	66	1	rep
DNA 螺旋酶Ⅱ	DNA 螺旋酶	82.1	1	uvrD
RNA 酶 H	切割引物 RNA，阻止非特异性启始	17.6	1	rnh
DNA 拓扑异构酶 I	消除负超螺旋	105	1	topA
Hu 蛋白	组织染色体结构	9.5	2	hupA.B.
蛋白质 n	组装 φ174 复制复合体	14	2	不清
蛋白质 n′	组装 φ174 复制复合体	76	1	不清
蛋白质 n″	组装 φ174 复制复合体	17	1	不清
DnaJ 蛋白	热激蛋白，与 DNA 结合	41.1	2	dnaJ
DnaK 蛋白	热激蛋白，5′－核苷酸磷酸酶	69.1	1	dnaK
GrpE 蛋白	热激蛋白，调节 DnaK 蛋白活性	24		grpE
GroES 蛋白	噬菌体组装，抑制 dnaA（T）突变	10.5	6～8	mopB（grocs）
GroEL 蛋白	噬菌体组装，抑制 dnaA（T）突变	60	14	mopA（groEL）
DNA 腺嘌呤甲基化酶	修复错误配对，调节 oriC 位点的复制启始	31	1	dam
（Arg－tRNA）				dnaY

（一）直接参与复制叉形成的蛋白质

1. 单链 DNA（ssDNA）结合蛋白（SSB）　SSB 为一种四聚体蛋白，分子量 18 873，大约可覆盖复制启始区中 33～65 个核苷酸的单链区。根据在离体条件下的研究，大肠杆菌 SSB 蛋白对其染色体复制具有多重功能。SSB 能够使双螺旋 DNA 松弛，但不能使其解螺旋。DNA 解螺旋由大肠杆菌的几种 DNA 螺旋酶（helicase）包括 DnaB 蛋白、Rep 酶和螺旋酶Ⅲ等所催化，但 SSB 与由螺旋酶催化形成的单链 DNA 结合则又极大地促进解螺旋过程。一般认为，SSB 在这一过程中的作用是阻止两条分开的单链再度形成双链。SSB 与单链 DNA 结合的目的也可能是使两条分开的单链免遭细胞中各种单链特异性的核酸酶切割。

在 DNA 复制过程中，SSB 还对其他一些 DNA 聚合酶如 DNA 聚合酶Ⅲ全酶的活性具有多重促进作用，其主要功能可能是使模板链保持牢固构型，不至形成次级结构。当然，SSB 与 DNA 聚合酶Ⅲ全酶直接作用至少可以提高 DNA 聚合酶的结合活性和合成活性。

SSB蛋白的生理功能不仅只限于促进复制叉形成过程，更重要的是在DNA复制启始之前，对在oriC位点形成核蛋白结构（包括DnaA、B、C和Hu蛋白复合体）并加以固定。另外，SSB蛋白通过包裹复制叉或其他位置解螺旋的DNA，消除细胞内引物系统的作用，提高DNA复制的特异性。

2．DnaB蛋白　DnaB蛋白为一种六聚体，分子量52.3kD，具有核苷三磷酸活性。在存在核糖核苷三磷酸的条件下，DnaB与DNA结合的能力较低，约覆盖80个碱基的区域，能够同DnaC蛋白结合，形成稳固的复合体。DnaB的基本功能是在复制机器之前使DNA解螺旋，它沿其中一条单链按5′→3′方向侵入双链部分。SSB、引物酶（primase）以及4种核苷三磷酸都能提高DnaB的螺旋酶活性。

DnaB蛋白除了在DNA合成机器之前将DNA解螺旋以外，它还在复制过程中具有第二种重要作用，即DnaG引物酶在合成RNA引物时，完全依赖于与模板DNA链结合的DnaB蛋白。一条与单链DNA结合的DnaB蛋白能使引物酶在一条模板链的许多不同位点合成多段引物RNA，可能DnaB使模板DNA链产生某种由DnaG识别的次级结构。

3．DnaG引物酶　DnaG引物酶为一条多肽链分子，分子量为65 600，为一种利福平抗性的RNA聚合酶，它以DNA为模板，合成互补的RNA链，其主要作用是在复制叉上合成启始后随链中冈崎片段的引物RNA。

由引物酶合成引物的过程还需要DnaB蛋白，在存在4种核苷三磷酸时，二者通过互作，可以在任何单链DNA分子上合成引物RNA。一般认为，引物酶具有提高DnaB蛋白与ssDNA形成的复合体的稳定程度。引物RNA长10～60个碱基，其长度受几种因素的影响。活体内的引物是一种脱氧核糖核苷酸和核糖核苷酸的混合物。引物中含有脱氧核糖核苷可能会使引物合成长度缩短，较短的引物可能易于被细胞内的核酸酶降解，其空缺便于由DNA聚合酶I合成DNA片段加以填补。

4．DNA聚合酶Ⅲ全酶　DNA聚合酶Ⅲ（DNA polymerase Ⅲ）全酶是大肠杆菌的一种复制型DNA聚合酶，全酶由10种不同亚基组成（表3－1），为一种非对称的二聚体，有2个催化中心，每一催化中心由一个α、ε和θ亚基组成。它除了具有按5′→3′方向合成DNA的聚合酶活性以外，还具有两种核酸酶活性，即3′→5′和5′→3′外切核酸酶的功能。其DNA聚合酶的功能主要由α大亚基执行，而3′→ 5′外切核酸酶活性由其ε亚基执行并需要单链DNA的3′－OH基末端。在DNA复制过程中，这种3′→5′外切核酸酶活性的作用是除掉增长链前端的非互补的核苷酸，提高DNA合成的精确性。DNA聚合酶Ⅲ催化核心中的5′→3′外切核酸酶活性需要5′单链末端，从5′端开始，逐渐降

解单链 DNA，直到 DNA 双链区为止，但在活体内并未测定到 DNA 聚合酶具有 5′→3′外切核酸酶活性。DNA 聚合酶Ⅲ全酶中其他辅助亚基如 τ、γ、δ′、X 和 ψ 具有提高催化核心活性的作用，β 亚基的二聚体加上一个 γ 复合体共同识别引物的模板链。其中一对 β 亚基形成所谓的夹板，结合到 DNA 外围。

5. DNA 聚合酶 I　DNA 聚合酶 I（DNA polymerase I）是大肠杆菌中发现最早的一种 DNA 聚合酶，只有一条多肽链，分子量为 103kD，在磷酸二酯键的形成和水解过程中行使三种功能。它具有典型的 5′→3′DNA 聚合酶活性，将单个碱基加到引物链的 3′－OH 基末端。同时，它还具有两种外切核酸酶活性，其中之一是 3′→5′校对外切核酸酶的作用，切割非配对碱基的 3′－OH 末端，形成 5′－脱氧单核苷酸（5′－deoxymononucleotid）。DNA 聚合酶 I 也具有 5′→3′外切核酸酶功能，但只从 5′端开始，降解碱基配对的 DNA，留下长度不超过 10bp 的寡聚脱氧核糖核苷酸。当 DNA 聚合酶 I 结合到双链 DNA 的某个切口或小缺口上时，可以使切口沿双链移动，这种反应叫做切口平移（nick translation）。在这一反应中，DNA 聚合酶 I 一方面进行 DNA 合成，同时也在增长的 DNA 链的前端按 5′→3′方向、以外切核酸酶的方式降解一条单链。

利用某些水解蛋白酶水解 DNA 聚合酶 I，可产生两条活性片段，较大的一条即羧基末端部分如克列诺片段（klenow fragment）分子量为 68kD，具有 DNA 聚合酶和 5′→3′外切核酸酶活性。较小的一条片段即 N 基末端部分分子量为 35kD，具有 5′→3′外切核酸酶活性。活体内，DNA 聚合酶 I 的切口平移作用对中断 DNA 复制过程极为重要，而 5′→3′外切核酸酶作用则是除掉新合成 DNA 片段的 5′端 RNA 引物。因此在活体内，DNA 聚合酶 I 一方面切割引物 RNA，同时也修补前引物所在位置的缺口。

6. DNA 连接酶　DNA 连接酶（DNA ligase）也只有一条多肽链，分子量为 75kD，催化两段相邻 DNA 片段的 3′－OH 和 5′磷酸基末端的连接反应，同时也在染色体 DNA 复制的延长期行使功能。

（二）在 OriC 位点启始 DNA 复制的蛋白质

1. DnaA 蛋白　DnaA 蛋白为一种具有复制子特异性的启始因子，同时也是细菌 DNA 复制所必需的蛋白质，它识别 oriC 的 DNA 序列，并同 oriC 进行特异结合。DnaA 蛋白的分子量为 52.5kD，为一条多肽链，它与 245bp 的 oriC 序列中多个位点进行特异结合，其中互作程度最高的为 4 段长 9bp 的重复序列，这 4 段序列叫做 DnaA 框，其共有序列为：

A　　A
TT　　A A　A
TT　C C
CC　　C G　C
C　　T

DnaA 蛋白与 oriC DNA 有很高亲和性，根据电镜观察，大约 20 至 30 条 DnaA 蛋白同时与 oriC DNA 互作，形成一个很大的核蛋白复合体结构，其中含有 200～250bp 的复制启始区 DNA。DnaA 蛋白的浓度是决定复制启始时间的关键，当它与 OriC 结合后，在 oriC 中长 13bp 的重复位点（富合 AT）首先将双螺旋打开，然后再通过 DnaC 将 DnaB 螺旋酶结合进来。

2. DnaC 蛋白　DnaC 蛋白为单分子，分子量 29kD，没有酶催化活性。但是，DnaC 蛋白能够与 DnaB 螺旋酶进行特异互作，形成一种稳定的 $(DnaB)_6$ － $(DnaC)_6$ 蛋白质复合体。DnaC 蛋白对 DnaB 蛋白的 ATP 酶活性不产生任何影响，但对由 DnaB 和引物酶所催化的引物合成反应具有促进作用，可能 DnaC 对稳定由 DnaB 螺旋酶与 ssDNA 结合形成的复合体具有重要作用。

DnaC 蛋白在大肠杆菌 DNA 复制过程中的确切功能尚有待确定，但一般认为其基本功能是将 DnaB 蛋白释放到 DNA 复制启始区。这一过程需要 $(DnaB)_6$－ $(DnaC)_6$ 复合体与 oriC 中 DnaA 蛋白进行特异互作。

3. DNA 回旋酶　DNA 回旋酶（DNA gyrase）催化几种反应，（1）催化依赖于 ATP 的共价闭合环状 DNA 的超螺旋化；（2）催化依赖于 DNA 的 ATP 水解作用；（3）在存在 ATP 的条件下，催化负超螺旋的松弛；（4）催化粘连或打结的双螺旋 DNA 的形成和溶解。在上述每一步骤中，DNA 回旋酶都须将 DNA 双链切断，然后再将双螺旋 DNA 的一条片段从切口处穿过，最后再将切口缝合。回旋酶交错切割 DNA，在 5′端形成 4 个碱基的单链尾巴。

DNA 回旋酶由两种多肽链组成，即回旋酶 A（97kD）和回旋酶 B（90kD），通常以四聚体（A2B2）存在。回旋酶 A 具有催化断裂—重接的功能，回旋酶 B 具有依赖于 DNA 的 ATP 酶活。

DNA 回旋酶在 DNA 复制过程中三个阶段即启始、延长和终止阶段起着关键作用。

复制延长期：环状双螺旋 DNA 分子在复制过程中需要一个旋轴将两条相互高度缠绕的单链分开，DNA 回旋酶则在这个过程中起着降低螺旋化程度的作用和直接制备复制叉的作用。DNA 回旋酶可能通过某种重要的酶催化机制直接作用于复制叉向前移动过程，如使平行的两条 DNA 链在复制叉处解螺

旋，或者使正在复制的染色体保持在解螺旋的负超螺旋状态。

复制启始：DNA 回旋酶趋向与 245bp 的最小 oriC 序列的右方某个位点结合。这种回旋酶－DNA 复合体中 DNA 长 140bp，其中还至少包括两个 DnaA 蛋白的识别位点。回旋酶还在 DNA 复制启始期起着某种间接作用，即对其他复制启始因子与 oriC 结合提供负超螺旋模板。

复制终止：许多研究表明，当环状 DNA 分子的复制完成以后，两条亲链之间还存在 10～30 个缠结点，这种缠结的二聚体可能是环状双螺旋 DNA 复制的一种中间产物，它可能是染色体末端未复制的亲本 DNA 不能与拓扑异构酶结合的结果。

DNA 回旋酶和拓扑异构酶 I 都能解除两个环状双螺旋 DNA 之间的粘着点，但后者至少需要两条双链 DNA 中各有一条单链有缺口。DNA 回旋酶参与将两条复制的子染色体分开的过程，而染色体分离过程则不一定需要拓扑异构酶 I。

（三）终止 DNA 复制过程的蛋白质

细菌 DNA 的复制终止过程可能需要数种不同的终止蛋白。在大肠杆菌中，染色体复制终止过程需要一种叫做 Tus 蛋白的终止因子。Tus 蛋白的分子量为 36kD，为可溶性单分子，可复盖 13bp 的 DNA 序列。Tus 蛋白与终止位点（terC）的 DNA 结合，形成阻止复制叉前进的 Tus－ter 复合体。枯草杆菌的复制终止过程则需要另一种分子量为 29kD 的终止蛋白（RtP），这种蛋白质通常以二聚体形式存在，能与复制终止位点中一对颠倒重复序列结合。

在染色体正常复制启始或复制延长期期间，还有一种调节蛋白也整合到 DNA 延长机器中，直到复制终止期这种蛋白质才行使功能，使复制复合体解体。这种调节蛋白就是 DnaT 蛋白，又叫 i 蛋白或 x 因子。

（四）其他蛋白质

DNA 复制过程还涉及许多其他蛋白质，如 DNA 螺旋酶（DNA－helicase）、RNA 酶 H（RNa－se H）、DNA 拓扑异构酶 I（DNA topoisomerase）、Hu 蛋白等。由于这些蛋白质对细菌 DNA 复制的确切作用目前还不清楚，所以通常就可把这些蛋白质统称为辅助复制蛋白，其主要功能见表 3－2，这里不作详述。

三、细菌 DNA 复制的模型

（一）大肠杆菌 oriC 的结构

大肠杆菌的染色体复制启始区（oriC）含有一组基因（图 3－3），这组基

因的转录可能与 DNA 复制有着某种关系。在这一区域中，能够使 DNA 自动复制的最小片段为 245bp。在其他一些革兰氏阴性细菌的染色体 DNA 复制启始区中也都存在相似的、结构高度保守的序列。大肠杆菌的最小 oriC 序列中存在 5 段 9bp 的结构高度保守的序列，DnaA 蛋白与这 5 个位点结合，所以这 5 个位点又叫 DnaA 框（DnaA box），分别以 R1、R2、R3、R4 和 R5 表示（图 3－4）。DnaA 框广泛分布于大肠杆菌基因组中，据分析平均每 10kb 的序列就含有 3.3 个 DnaA 框。在革兰氏阳性细菌如枯草杆菌基因组中也存在多个 DnaA 框，这种现象表明，DnaA 框在革兰氏阳性和阴性细菌中都是高度保守的。

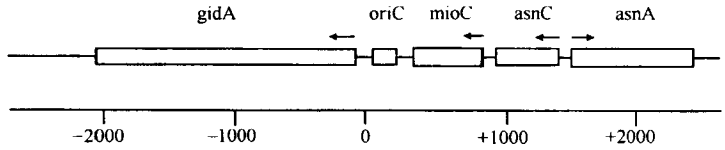


图 3－3　大肠杆菌 oriC 区的遗传图

上部箭头示转录方向，下方数字为从 +1 起的碱基数

大肠杆菌最小 oriC 序列的另一特征是其中存在许多 GATC 位点，这种位点是 Dam 甲基化酶作用的目标位点，所以叫做 dam 位点。在其他革兰氏阳性和阴性细菌中也都存在这种结构。一般认为，oriC 中这些 GATC 位点可能对 oriC 中形成 DnaB/DnaC 蛋白复合体起着某种作用。

在 oriC 中还存在与膜蛋白结合的位点，可能 oriC 与膜蛋白结合对其本身附着到质膜上具有重要作用，或者对活体内 DNA 从单向复制转向双向复制起着某种作用。

（二）细菌染色体 DNA 的复制

根据对含有大肠杆菌 oriC 的质粒复制的研究，染色体 DNA 复制可以分成几个不同阶段（图 3－4）。首先是 DnaA 蛋白结合到 oriC 中各个识别位点上，形成一种特殊的核糖核蛋白结构，以用作 DnaB 螺旋酶转移到模板 DNA 上的座位。然后，DnaB 螺旋酶将 oriC 中双螺旋解螺旋，同时 SSB 蛋白与解螺旋的 DNA 结合并促使引物酶结合进来。当引物 RNA 合成以后，DNA 聚合酶Ⅲ全酶在 DNA 回旋酶帮助下，迅速沿模板链前进，延长 DNA 复制链。其具体步骤分述如下。

1. *在 oriC 组装和活化 DNA 复制机器*　关于大肠杆菌从 oriC 开始的双向复制如图 3－4 所示。在这个模型中，细菌染色体 DNA 复制启始时，许多条 DnaA 蛋白分子与 oriC 中各个 DnaA 框结合，其他一些复制启始因子则直接通过蛋白质－蛋白质互作与 DnaA 蛋白结合，在复制启始区形成类似核小体（nucl－eosome）的结构。当存在 ATP 时，这种特化的核糖核蛋白复合体就具

有特殊的、具有活性的构型（图 3－4b）。

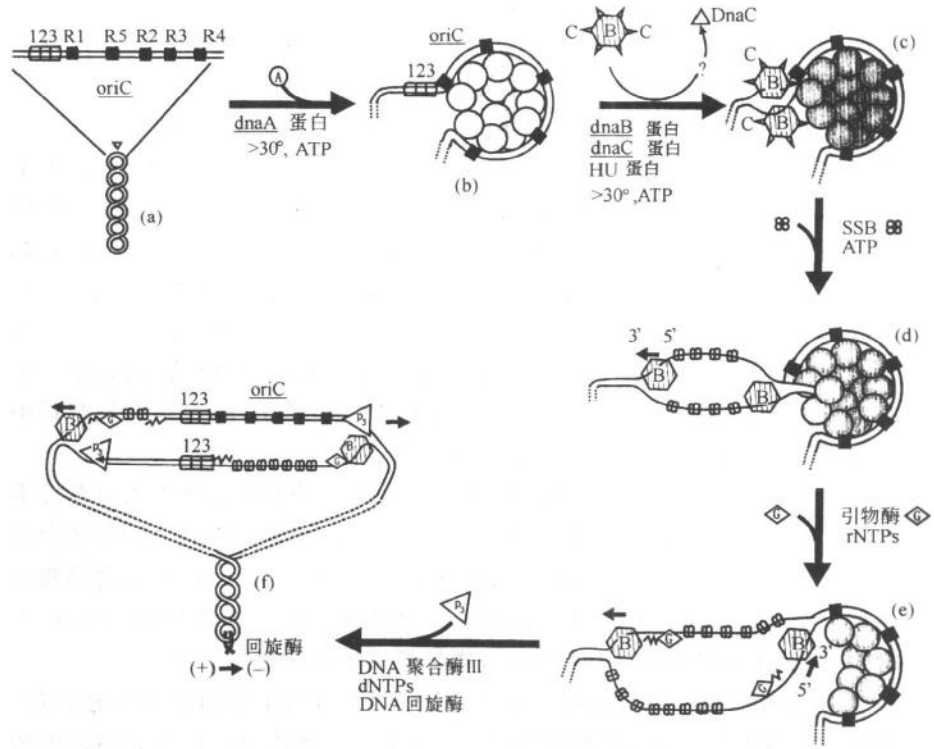


图 3－4　大肠杆菌染色体 DNA 的双向复制模型

R1 到 R5 为大肠杆菌 DNA 蛋白的识别序列，未涂黑的方块 1、2、3 为 DnaB－DnaC 蛋白复合体的识别位点（根据 MaCken 等，1987）

如果模板 DNA 为负超螺旋，则 oriC－DnaA 蛋白复合体就作为 DnaB 螺旋酶与细菌复制启始区结合的座位，形成所谓的前原引物复合体（prepropriming complex）（图 3－4c）。这一步骤完全依赖于 DnaB 和 DnaC 蛋白之间形成紧凑的复合体。形成的前原复合体含有一条 DnaB 螺旋酶分子，其中 DnaB 结合到与 DnaA 复合体并列的每一条 oriC 链上（图 3－4c）。

当 DnaB 蛋白处于适当位置以后，一条 DnaB 螺旋酶分子则从 DNA 双链之间穿过，沿着它结合的链，按 5′→3′方向离开 DnaA 复合体，启始 oriC 左方解螺旋的 DNA 复制（图 3－4d）。同时 DnaB 螺旋酶还使 SSB 蛋白与 ssDNA 结合。位于 oriC 上其他 DnaB 分子则开始向右移动。

在 DNA 双螺旋开始解螺旋后不久，DnaG 引物酶识别每一复制叉上的 DnaB 螺旋酶，并通过与 DnaB 互作，合成引物 RNA。同时，DnaA 复合体脱

离 oriC，而向右移动的 DnaB 螺旋酶则穿过复制启始区。在大肠杆菌中，引物合成涉及约 20 种蛋白质，它们同 DNA 模板链结合，形成所谓的引物体（primosome）。DnaG 引物酶是其成分之一，当引物体沿后随链的模板链向前移动时，引物酶则以相反方向即按 5′→3′方向合成引物。

从图 3－4 可以看出，向右延长新链的第一个引物是在 oriC 左方位点上合成的，而向左延长链的第一个引物是在 DnaA 盖合 R1 附近某个位点，通过引物酶－螺旋酶互作合成的。这些最初引物然后再由 DNA 聚合酶Ⅲ全酶延长，成为前导链（图 3－4f）。在后随链的每一个复制叉处，通过 DnaB 对 DNA 双螺旋连续不断地解螺旋，因此就能连续合成前导链和通过引物酶合成后随链的多条中间产物即冈崎片段。在这个模型中，前导链和后随链的所有引物都是由 DnaG 引物酶合成的。实际上，通过 DnaB 螺旋酶和 DnaG 引物酶互作最先合成的各条引物都是用于两条方向相反的前导链的合成（图 3－4e、f）。

有关从 oriC 开始的、DNA 双向复制的启始及复制叉延长过程还存在许多有待澄清的问题。例如上述复制模型忽略了在 DNA 复制过程中 RNA 聚合酶的作用。另一重要问题是在复制启始过程中，有关 DnaC 蛋白对形成前引物复合体（图 3－4c）的确切作用还不清楚，可能 DnaC 蛋白的作用是与 oriC 上 DnaA 框互作，并通过这种互作将 DnaB 螺旋酶置于复制启始区中。

2. 复制叉的延长　当引物体合成了引物之后，DNA 复制便进入延长期。有关复制叉的结构及其延长过程如图 3－5 所示。一般认为，启始后随链中冈崎片段合成有关的复制装置比较简单，由 4 种主要蛋白质组成，即 DnaB 螺旋酶、DnaG 引物酶、SSB 和 DNA 聚合酶Ⅲ全酶。DNA 聚合酶Ⅲ全酶是延长过程中一种关键酶，它同引物体结合，形成所谓的复制体（replisome），执行 DNA 复制的延长过程。DnaB 蛋白一方面行使螺旋酶的功能，沿着后随链的模板链，按 5′→3′方向逐渐向前移动，不断地将双链 DNA 解螺旋，使 DNA 聚合酶Ⅲ不断延长前导链。另一方面，DnaB 蛋白也同 DnaG 引物酶一起参与合成后随链中、形成冈崎片段所需的各条引物 RNA。

那么，复制体又是如何并列合成前导链和后随链呢？柯恩伯格（A.kornberg 1992）认为，后随链的模板链环出复制体（图 3－5），使复制体能够同时阅读两条单链。前导链的模板链从上部进入复制体，后随链的模板链则从下部进入复制体而形成环状结构（图 3－6）。这一假说认为复制体是一个不对称的二聚体，其中一个成员合成前导链，另一成员合成后随链。毫无疑问，这种假定与 DNA 聚合酶Ⅲ全酶有 2 个活性中心的事实相符。这种模型的最大优点是一个复制体能够连续复制两条链而无须离开复制叉和再度与 DNA 结合。

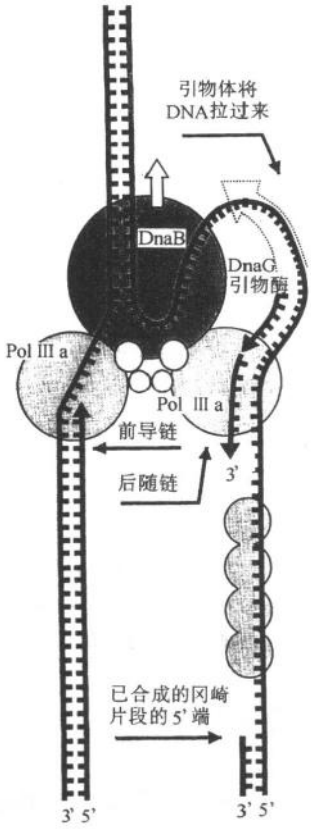


图 3－5　DNA 聚合酶Ⅲ的两个催化核心分别合成 2 条子链

DnaB 使复制叉前移，引物体将后随链的模板链拉过复制体

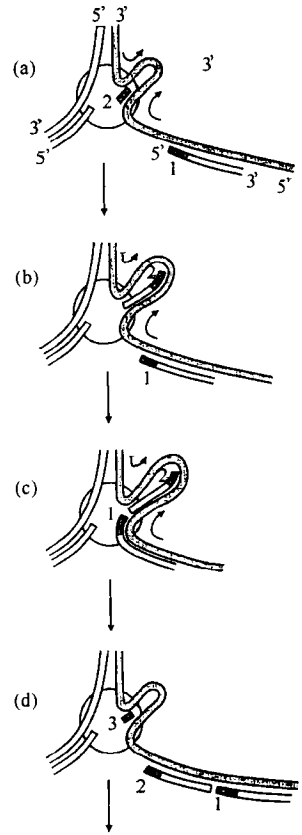


图 3－6　同时合成两条 DNA 单链的模型

A. 后随链的模板链通过复制体形成环，引物酶合成引物　B. 复制体从上到下阅读后随链，环扩大，环的下部分使第 2 个冈崎片段延长　C. 第 2 个冈崎片段进一步延长，使其一端与第 1 个冈崎片段的引物相邻　D. 复制体释放出环、冈崎片段 1 和 2，并形成一个新的环，使引物酶合成新的引物，开始下一轮冈崎片段合成

3. DNA 复制的终止　在大肠杆菌和枯草杆菌中，在 OriC 中形成的两个复制叉沿环状染色体双向移动，最后在与 oriC 相对的复制终止子 ter 终止。大肠杆菌染色体上有 6 个复制终止位点（图 3－6）。其中有 3 个终止子信号终止向反时针方向延伸的复制叉，另外 3 个则终止向顺时针方向延伸的复制叉。各段 ter 序列长 22－23bp，其共有序列为

AA　AA　　A
NN　G　TGTTGTAACTA　NNN
TT　TG　　C

但其中真正不变的核苷酸只有 11 个。在大肠杆菌染色体的复制终止过程中复制叉前进到终止位点时，Tus 蛋白与 terC 位点进行特异结合后，阻止复制叉前进。枯草杆菌染色体只有一个复制终止信号，由两段相距 59bp 的颠倒重复序列组成，当 Rtp 复制终止蛋白与这两段颠倒重复序列结合后，终止复制过程。

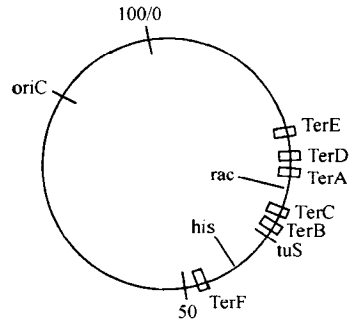


图 3－7　大肠杆菌染色体终止位点（TerA－TerF）在染色体上的相对位置

终止位点 TerA、TerD 和 TerE 终止反时针方向前进 的复制叉，而 TerB、TerC 和 TerF 则终止顺时针方向前进的复制叉；tus 为编码 Ter 结合蛋白的基因

四、质粒的一般特征及其复制调节

细菌细胞除了有一个大的双链环状 DNA 以外，还有一至数个小环状的双链 DNA，叫做质粒。质粒只载有少数基因，如抗生素抗性基因等，其复制不依赖于主染色体，能够独立自主地复制。质粒在细胞中的拷贝数是稳定的，即其数目在每一个细胞世代保持不变。还有一些质粒既可以自主的染色体外因子存在，也可以插入到细菌染色体上，作为染色体的一部分进行传递，这种类型的质粒叫做附加体（episome）。某些质粒和附加体可以通过细菌接合（conjugation）在细胞之间进行转移。在接合过程中，有时宿主基因也随质粒从一个细胞转移到另一细菌细胞，从而使遗传信息在细菌之间进行交换。

根据细胞中质粒的拷贝数，可将质粒分成两大类。第一类为单拷贝质粒（single copy plasmid），这类质粒在每个细胞世代只复制一次，即质粒中唯一的一个复制启区（oriP）只复制一次，然后子复制启始区分离到两个不同子细胞中。因此每一细胞中的染色体和质粒都只有一个拷贝。第二类为多拷贝质粒（multicopy plasmid），这类质粒在每个细胞世代中可以进行多次复制，所以每个细菌细胞可以有多个拷贝的质粒。有些质粒能够产生一些毒素和解毒素，以保证细菌细胞必须含有质粒。毒素较稳定，解毒素寿命较短，具有抑制毒素的

作用。当细菌细胞丢失质粒后，解毒素降解，然后毒素引起细胞死亡。所以丢失了这类质粒的细菌必定死亡。这种杀伤质粒有两类，第一类可以 F 质粒为代表，它们编码的杀伤因子和解毒素都是蛋白质，解毒素解除杀伤因子的作用。第二类可以 R 1 质粒为代表，它们编码的杀伤因子为毒素蛋白，而解毒素则为一种反义 RNA（antisenseRNA），反义 RNA 通过与毒素 mRNA 进行碱基互补配对，阻止毒素 mRNA 翻译。

细胞中质粒的拷贝数还与质粒的不相容性（imcompatibility）有关。质粒不相容性是指在一组质粒中，其成员不能共存于同一细菌细胞中的现象。其主要原因是这些质粒在 DNA 复制和分离阶段不能相互区别。如果将属于同一相容群的具有不同复制启始区的两个质粒引入同一细胞中，这两种质粒都不能复制，因为同一细胞已有 2 个复制启始区。当经过细胞分裂，两个质粒被分离到不同子细胞中后，这两种质粒才能在不同细胞中复制。另一方面，如果这两个质粒都有相同的决定分离的位点，那么这两个位点之间就会通过竞争，使两个质粒被分离到不同细胞中，而不能共存于同一细胞中。

细胞中质粒拷贝数是由质粒的复制启始过程决定的。例如大肠杆菌的 ColE 1 质粒在每个细菌细胞中大约有 20 个拷贝，保持这种拷贝数取决于 E-olE1 复制启始区中启始复制的机制。

质粒 DNA 复制时，首先在复制启始区上游转录出一段长 555 个核苷酸的 RNA，当转录进行到复制启始区时，RNA 酶 H 在复制启始区中切割转录子，产生 3′－OH 末端。然后 DNA 聚合酶再以这段 RNA 为引物，开始合成 DNA。

形成 RNA 引物受两种因素的影响。其中之一是受合成一段与引物 RNA 互补的 RNA 如 RNAI 的影响。另一种因素是受由附近座位编码的蛋白质的影响。RNAI 由引物 RNA 链的非模板链编码，长 108 个碱基（图 3－8）。RNAI 的合成从引物区开始，在引物 RNA 合成开始的位点附近终止。因此 RNAI 与引物 RNA 的 5′区互补。这两种 RNA 通过碱基配对控制引物 RNA 参与 DNA 复制过程。RNAI 叫做逆转录子（countertranscript），又叫做反义 RNA。与前体引物 RNA 碱基配对后前引物就形成一种特有的次级结构，阻止 RNA 酶 H 在复制启始区内切割，从而不能为 DNA 复制提供引物。在这个过程中，前体

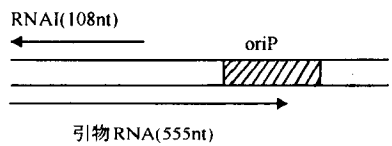


图 3－8　RNAI 与引物 RNA5′区域互补

引物 RNA 作为启始复制的正调节子，而 RNAI 则为一种负调节子，抑制正调节子的作用。RNAI 与引物 RNA 的结合过程受一种叫做 Rom 蛋白质的影响。Rom 蛋白由复制启始区下游的一个基因编码，它加强 RNAI 与前体引物 RNA 结合，从而阻止引物形成。上述现象表明，质粒通过复制启始过程来控制其在细胞中的拷贝数。自然的细菌质粒通过人工改造，可以用作克隆基因的载体，即通过离体重组，将某个基因或一段 DNA 序列连接到质粒上，然后通过转化细菌，将其重新送回到细胞中（见第九章）。

第二节　细菌基因组的转录与翻译

一、与转录有关的酶

（一）依赖于 DNA 的 RNA 聚合酶

转录需要 RNA 聚合酶，在原核生物中只存在一种 RNA 聚合酶，转录所有基因。在真细菌中，RNA 聚合酶由 4 到 5 个亚基组成，其活性受到利福平（rifampicin）和利迪链菌素（streptolydigin）的抑制。大肠杆菌 RNA 聚合酶的分子量为 440kD，由 4 个亚基即 $\beta'\beta\sigma\alpha$ 组成，各个亚基的分子量分别为 160kD、150kD、70kD 和 40kD。α 亚基是 RNA 聚合酶的核心酶，它催化 RNA 链延长，但它缺乏启始特异性。与转录启始过程有特异性的亚基为 σ 亚基，σ 亚基较易从全酶上分离下来，形成 $\beta'\beta\alpha 2$ 四聚体。σ 亚基从全酶上分离与 RNA 聚合酶选择转录位点有关，大肠杆菌中 σ^{70} 为一种普通亚基，与大多数基因转录有关，但还存在其他基因特异性的 σ 亚基（表 3－3）。β′是其中唯一能与 DNA 和 σ 亚基结合的亚基。在枯草杆菌中，RNA 聚合酶有 2 种 σ 亚基，即 σ^{55} 和 σ^{28}，其分子量分别为 55kD 和 28kD。σ^{55} 的功能是分别与启动子中－35 位的－TTGACA－和－10 位的－TATAATA－框互作。此外 σ 亚基还可以同转录起点上游第 161

表 3－3　大肠杆菌不同 σ 因子及其识别序列

σ 因子	识别的启动子	启动子同感序列	
		－35 区	－10 区
σ70	大多数基因	TTGACA	TATAAT
σ32	热激效应诱导的基因	TCTCNCCCTTGAA	CCCCATNTA
σ28	游动性和趋化性基因	CTAAA	CCGATAT
σ38	生长稳定基因	不清	不清
		－24 区	－12 区
σ54	氮素代谢及其他功能的基因	CTGGNA	TTGCA

至 186 之间的一段序列互作。

在某些高级细菌（如 *Salfolobus acidocaldaris*）中，RNA 聚合酶的结构比真细菌 RNA 聚合酶复杂得多，由 2 个大亚基和 8 个不同小亚基组成。两个大亚基的分子量变动于 44 000 至 10 800 之间，全酶中小亚基总分子量为 423 000。*Salfolobus acidocaldaris* 中 RNA 聚合酶的转录活性受放线菌素（actinomycin）和肝素（heparin）的抑制，但不受利福平、利迪链菌素和 α－鹅膏蕈碱（α－amanitine）的影响，这种抗性特征与真核细胞的 RNA 聚合酶 I 相似（见第四章）。

（二）CAP 因子及其识别位点

CAP 因子是一种与 cAMP 结合的蛋白质，其功能是控制大肠杆菌中某些操纵元的表达，如 lac 和 ara 操纵元等（本章第三节），它与确定细胞内 cAMP 的水平有关。当细胞中 CAP 因子的浓度较高时，能使某些操纵元活化。CAP 因子由 2 个亚基组成，总分子量为 50 000，只有在与 cAMP 形成复合体后才能结合到 DNA 上。当 CAP 因子与 lac 操纵元编码区上游某个位点结合时，可使该操纵元的转录水平提高 50 倍，其识别位点的序列为 GTGAT－TA－CT-CAC。通常认为，CAP 因子、σ 因子以及核心 RNA 聚合酶三者在 DNA 分子上相连成一线，各自覆盖各自的识别位点。但是，在 ara 操纵元的活化过程中，除了需要 CAP 因子和完整的 RNA 聚合酶外，还需要一种 C 蛋白，而在革兰氏阳性细菌枯草杆菌中，尚未发现 CAP 因子。

（三）其他辅助因子

在大肠杆菌中，有几种因子能提高基因的转录活性。第一种为 H 因子，已知其中 H1 和 H2 影响 RNA 聚合酶的活性，二者都是低分子量的蛋白质。H1 为寡聚体，中性；H2 为单分子，强碱性。H1 和 H2 因子结合在一起时，其增效作用比单独存在时高许多倍。

另一辅助因子为 Hu（Hu factor），其热稳定性较高，分子量为 7 000，它与 H1 的差别在于 H1 刺激基因转录的效率因基因不同差别很大，而 Hu 的活性功能则不具有特异性。第三个转录辅助因子叫做 D 因子（D factor），其特性目前还不很清楚，只知它能提高由大肠杆菌 RNA 聚合酶催化的 λ 噬菌体基因的转录。

（四）核糖体对转录的作用

在许多细菌中，当编码蛋白质的基因开始转录以后，核糖体必须迅速结合到正在延长的 RNA 链上，否则转录过程就不能正常进行，就不能转录出成熟的 mRNA。

二、转录启始、延长与终止

（一）启动子的结构特征

转录启始过程相当复杂，包括 RNA 聚合酶结合以及这种复合体的异构化，形成稳定的开放复合体和转录启始等过程。在活体条件下，RNA 启始合成的位点在基因转录起点上游，RNA 聚合酶中的 σ 单位与这一位点结合，形成启始复合体，所以叫做启动子。启动子中含有两段结构保守序列，一段位于转录起点上游第 10 个碱基的位置，即 -10 位，叫做普里布诺框，其共有序列为 TATAAAT；第二段位于这一位点上游 25 个碱基的位置即 -35 位，通常称为 RNA 聚合酶识别位点，其共有序列为 TTGACA，它是 RNA 聚合酶与启动子互作的最初位点。普里布诺框与 RNA 聚合酶识别位点之间的距离对启动子启始转录过程的效率影响极大。一般认为，二者之间相距 17 个碱基时，启动子具有最高效率。在不同细菌以及同一细菌的不同操纵元中，启动子中这两个位点的结构保守序列与上述共有序列略有差别，但它们都能很好地行使功能。如果这两个位点中发生点突变，可提高或降低启动子的效率（图 3-9）。因此，这两个位点的序列结构特征对确定启动子的效率具有重要作用，可能具有弱启动子的基因转录频率较低，而含有强启动子的基因转录频率较高。有些细菌如枯草杆菌的这两个位点的序列非常相似，但大肠杆菌的启动子在枯草杆菌中不能行使功能，而枯草杆菌的启动子在大肠杆菌中则具有功能，可能枯草杆菌基因在表达时使用不同的启动子序列，因为其细胞中存在几种不同的 σ 因子，

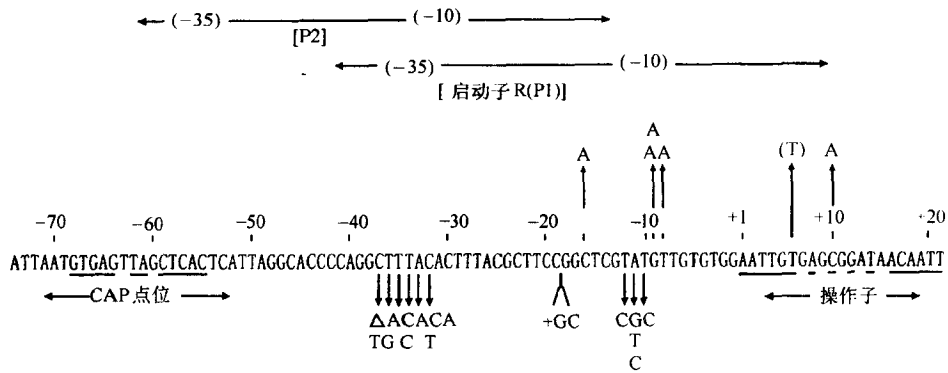


图 3-9　lac 操纵元的启动子和操作子

+1 为 mRNA 转录起点，正数为其下游序列，负数为上游序列；影响 lac 启动子活性的突变包括加强 lacP 功能的（线上碱基）和降低 lacP 功能的（线下碱基）位点；lac 操作子中有 6 个碱基的对称结构；P1 和 P2 为两个重叠的启动子

不同 σ 因子识别不同 DNA 序列。

（二）转录启始

根据对大肠杆菌转录启始过程的研究，有些大肠杆菌启动子每 5～10s 就启始转录一次（如 rRNA 基因），而有些启动子在一个细胞世代中才启始转录一次（如 lacI），因此启始频率或启动子的功能可相差 1 000 倍，甚至 10 000 倍。由于原核细胞中只有一种 RNA 聚合酶参与转录过程，说明细菌基因的转录启始频率直接与启动子序列的结构差异有关。可能这种差异反应到某个启动子序列与同感序列的同源程度方面，如果同源程度低，则启始频率低。对生活的细胞来说，启动子的最佳功能并不等于最高功能，所以有些启动子的实际启始频率很高，有些则很低。

通常认为，启动过程包括两个基本步骤：（1）RNA 聚合酶与启动子结合，形成所谓的关闭复合体（closed complex），其中的 DNA 为双链构型。（2）然后关闭复合体转变成用于启始 RNA 合成的开放复合体，即从 -10 位到转录起点之间的 DNA 双链分开成两条单链，这种 RNA 聚合酶与启动子互作的模型可用下列公式表示：

$$R + P \underset{KB}{\rightleftharpoons} RPc \xrightarrow{K_2} RPo \xrightarrow{\text{四种 NTP}} \cdots\cdots \rightarrow RNA$$

其中 KB 为结合常数，RPc 为关闭复合体，RPo 为开放复合体，K_2 为从 RPc 转变到 RPo 的速率常数。关闭复合体形成以后，启动子区局部熔解，形成所谓的转录泡（transcription bubble），熔解区的大小及范围因启动子不同而有差别，平均长 12bp，即从 -9 位至 +3 位之间的区域。开放复合体形成以后，RNA 聚合酶以其中一条单链为模板，按 5′到 3′方向合成 RNA。随着转录进行，σ 亚基从复合体上脱离下来，与另一条核心酶结合，参与启始其他基因的转录。在大肠杆菌中，σ 亚基在启动子上的释放位点目前还不清楚，但通常认为 σ 亚基释放之时标志着转录启始期的结束。

当 σ 亚基从转录启始复合体上释放下来之时，往往发生流产转录启始现象，即不能合成正常长度的 RNA 链。即使有足量核苷三磷酸存在，RNA 聚合酶也只合成一段短的、长 2～9 个核苷酸的序列，并将其从复合体中释放下来，重新启始转录过程。转录的这一阶段叫做流产启始期（abortive inituation）。流产启始转录可能是一种转录调节机制。

细菌中有些基因的转录启始除需要 RNA 聚合酶外，还需要在某些活化蛋白的帮助下，启动子才能有效地启始转录过程。但也有一些基因的启动子在无

活化因子的条件下，也能有效地启始转录过程。

转录活化子的作用机制主要有两种。第一种是活化子（如 CAP 蛋白）帮助 RNA 聚合酶结合到启动子上和促使闭合转录复合体转变成开放复合体，即所谓的补给模型（recruitment model）。另一类活化子（如 NTRC 蛋白）只作用于 RNA 聚合酶与启动子结合之后，它们促进闭合转录复合体转变成开放转录复合体的异构化过程。

转录启始过程还受 DNA 超螺旋的影响。DNA 超螺旋影响许多启动子的转录启始过程，在某些情况下，超螺旋提高基因的转录水平，而在另一些情况下，则降低基因的转录水平。在大肠杆菌中，超螺旋的形成与 DNA 回旋酶（DNAgyrase）基因（gyrA 和 gyrB）有关，而这两个基因的表达又受到 DNA 螺旋化程度的调节。

DNA 甲基化（DNA methylation）是改变转录启始频率的另一机制。由于转录启始涉及对启动子 DNA 序列的准确识别，当启动子的碱基发生修饰后，转录启始频率就会发生变化。碱基修饰的一个共同特征是，GATC 序列中形成 N6－甲基腺嘌呤。在大肠杆菌中，这一过程由 dam 腺嘌呤甲基化酶催化。例如 Tn5 转座酶启动子的序列为 GATCTGATCTTCCA，其中含有两个 dam 位点，当这两个位点完全不甲基化和完全甲基化时，前者转座酶 mRNA 的合成水平比后者高出许多倍。

转录启始过程还受许多其他因子的调节，特别是那些编码分解代谢功能的操纵元，当细胞处于与某种代谢途径有关的基质中时，这类操纵元才能表达。对于一些编码合成代谢的操纵元来说，当细胞处于与某种代谢途径有关的末端产物之中时，这类操纵元的表达水平则很低。有关各类操纵元的调节表达将在本章第三节以及以后有关章节中加以讨论。

（三）转录延长

当新合成的转录子增长到 8～9 个核苷酸之后，σ 因子脱离复合体，转录泡达到 18bp 长，即熔解的 DNA 和 RNA－DNA 杂种分子所在区域达到成熟，此时转录进入延长期。RNA 链延长既是一种由酶催化的多聚化反应，同时也是一种拓扑学反应（topological reaction）。因为这一反应一方面需要 RNA 聚合酶根据模板 DNA 的碱基顺序将四种活化的核苷三磷酸聚合在一起，同时也需要将 DNA 双螺旋不断解螺旋和再螺旋化。也就是说，RNA 聚合酶沿 DNA 移动是一种局部熔解的 DNA 泡在 DNA 分子内移动的过程。

当 RNA 链增长到 8～9 个核苷酸后，转录过程从流产启始期转变到延长期，转录复合体会发生一系列显著变化，最后才使 RNA 链延长急剧加快并延续下去。

第一，当σ因子脱离复合体后，可能提高RNA聚合酶与其他调节蛋白结合的能力，从而促进转录。

第二，RNA聚合酶的构型发生改变，同时与DNA模板链和非模板链接触的特性发生急剧变化。在转录延长期，RNA聚合酶对一些水解蛋白酶的抗性提高，同时它也使DNA对DNA酶I的敏感性发生改变。

第三，发生拓扑学变化。在发生上述变化的同时，RNA－DNA杂种分子增长到12bp，约相当A－型双螺的一圈，并以A－型核酸构型存在。此外，DNA泡的大小也由12bp增加到18bp，即达到成熟转录泡的大小。

第四，转录复合体稳定性增加。RNA聚合酶具有一个在转录启始期与模板DNA结合的结合沟（binding groove），在转录延长期，这个沟也与RNA－DNA杂种结合（图3－10）。如果RNA聚合酶与杂种分子的亲合力增加，则转录复合体在延长期的稳定性就会提高。一般认为，转录延长过程是一种吸收单核苷酸的循环过程。一个循环包括如下几个步骤：

1．RNA－DNA杂种分子在12＋1位增加一对碱基。

2．杂种分子在12－1位分开一对碱基，使杂种分子始终保持12bp长。

3．DNA双链在转录泡前端打开一对碱基。

4．RNA聚合酶向前移位一对碱基，转录产物的末端位点和基质位点朝模板链下游移动一对碱基。

5．DNA双链上在转录泡后形成一对碱基对。

6．新合成的RNA转录子伸入基质中一个碱基。

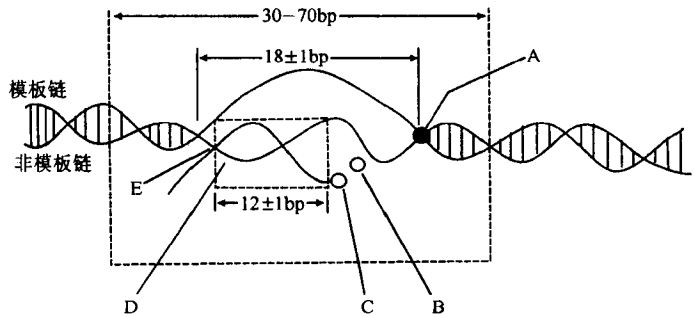


图3－10　转录泡的结构

其中RNA聚合酶覆盖的DNA区在转录启始期为70bp，在延长期约为30bp；B－基质，C－转录产品的末端位点，D－核酸结合沟，E－为RNA－DNA杂种分离位点，A－固定DNA泡前沿的未知因子，T和NT分别为模板链和非模板链（根据Yager和Hippel，1987）

虽然转录延长是RNA聚合酶活动的一种极稳定的方式，但在某些情况下，延长过程可以临时或永久性地被中断。根据离体条件下的研究，DNA模

板链的完整程度，RNA 聚合酶和各种 NTP 的纯度、反应基质浓度过高或过低，DNA 转录泡以及 RAN－DNA 杂种分子的稳定性等都会影响 RNA 链的延长过程。

转录延长期间，新合成的 RNA 链相当快速地从模板 DNA 链上脱离下来，分开的两条单链 DNA 又重新形成双链分子。大肠杆菌中，RNA 链的增长速度大约是每秒钟 60 个核苷酸，其合成速率比 DNA 复制略低。

（四）转录终止

转录过程也在 DNA 模板链的特殊位点上终止，但是终止信号一般位于 RNA 链上而不是 DNA 链上。原核生物的 RNA 聚合酶对两套不同的转录终止信号作出反应。第一种为内向终止子。具有一串碱基 A，在这串碱基上游有一对能够形成发夹结构的颠倒重复，其主杆中含有高比例的 G－C 碱基对。一般情况下，当 RNA 聚合酶遇到终止信号时，转录过程就自行终止，其中最典型的例子是大肠杆菌的 rRNA 的转录，其终止过程不需要其他酶类。但是，大肠杆菌中另一些基因的转录则需要第二种蛋白质即 ρ 因子（rho factor）参与，这种终止子叫做依赖于 ρ 因子的终止子。如果缺少 ρ 因子，RNA 聚合酶阅读 DNA 序列时就会通过终止信号，转录该位点下游的基因。ρ 调节的转录终止过程是由结合到新合成的转录子上的 ρ蛋白与停留在下游位点上的 RNA 聚合酶的互作而引起的。ρ 结合位点通常 C 较多而 G 较少，不具有序列保守性。

目前已知的转录终止控制机制主要有两类，第一种机制是通过转录复合体的变化，或者提高终止效率，或者降低终止效率。它可以通过某种蛋白质与转录子和 RNA 聚合酶结合，将正在延长的转录复合体修改成一种对转录终止信号作出反抗的形式，即所谓的反终止子（antiterminator），使转录继续进行。也可以是某种蛋白质与转录复合体互作，促进终止作用而不是反终止作用。第二种控制机制是 RNA 构型在终止子和反终止子之间互变，它与转录子中特殊序列和特殊结构有关，目前只见于内向终止子系统之中。有关转录终止和反终止的详细过程将在有关章节中分别进行讨论。

（五）细菌中 RNA 合成与蛋白质合成的偶联

在同一操纵元中，一个基因的多肽链终止密码对下游基因的转录具有重要影响。即使这些基因的编码序列并不由任何终止密码隔断，其转录效率也非常低。这种上游基因的终止密码干扰下游基因转录的现象叫做遗传极性（genetic polarity）。它是由于翻译失败而终止下游 RNA 合成的一种现象，细胞中与这一过程直接有关的因素是 ρ 上游顺反子转录终止因子。一般情况下，核糖体结合到 RNA 上，就可使 ρ 因子与正在合成的 RNA 结合，阻止 RNA 聚合酶的作

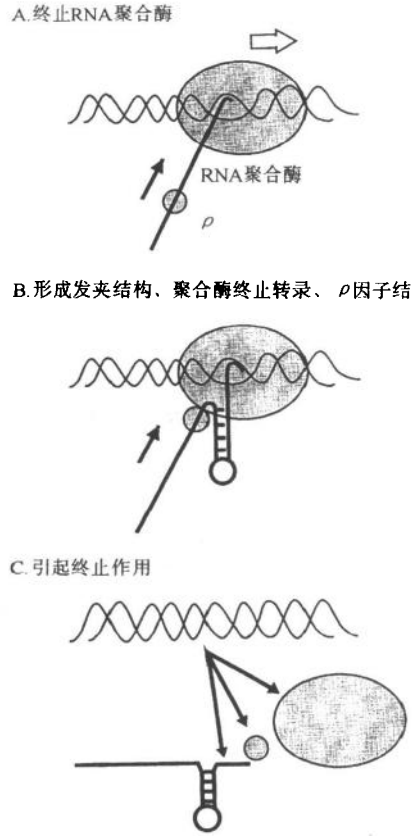


图 3－11　依赖于 ρ 因子的终止作用

A. RNA 聚合酶合成 RNA 时，ρ 因子与转录子结合，中止 RNA 聚合酶的作用　B. 转录子形成发夹结构，聚合酶中止转录，ρ 因子结合进来　C. ρ 因子解除 RNA－DNA 的氢键并释放转录子

用。但是当核糖体从终止密码上脱离后，RNA 就不能与 ρ 因子结合，转录就能继续进行。显然，终止密码与某个新核糖体结合位点之间的距离越小，ρ 因子作用的可能性就越大，极性表现也就更为严重。一般认为，极性代表细胞内一种阻止无效转录的机制，当细胞受饥饿，不能维持其有效蛋白质合成时，就会表现出极性现象。

当新合成的 RNA 未与核糖体结合或含有次级结构时，ρ 因子也能结合进来，但它并不能在 DNA 分子的任意位点阻止 RNA 聚合酶前进，ρ 因子只是在 RNA 聚合酶中止转录的 DNA 位点上释放 RNA。因此，RNA 合成并不停止在终止密码上，而是停留在翻译终止密码下游、第一个 RNA 转录终止子的位点上。

（六）RNA 在转录后的加工

细菌中有些 RNA 在转录后就可直接参与蛋白质合成，如某些 tRNA 和 rRNA，有些 mRNA 在转录尚未结束时就已开始蛋白质合成。但是有些初级转录子却含有几种不同的 RNA 分子，如大肠杆菌基因组中有 7 个 rRNA 操纵元，每一个都编码 3 种不同的 rRNA 和某些 tRNA，所以这些操纵元的转录子必须通过切割、加工，才形成成熟的 RNA 分子，最后参加蛋白质合成过程。

在大肠杆菌中，各个 rRNA 操纵元的转录子由 RNA 酶Ⅲ在特殊位点上水解磷酸二酯键，将各种 RNA 分子切开。rRNA 初级转录子在某些区域可以通过分子内碱基配对形成次级结构（图 3－12），使成熟 16S 和 23SrRNA 分子在双链区的末端环出，RNA 酶Ⅲ则作用于双链区，交错切割双链部分。然后，各条 tRNA 的末端进一步通过剪修，最后形成各种成熟 RNA。

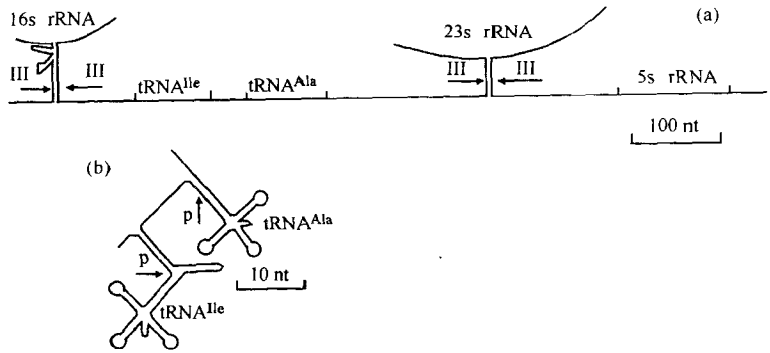


图 3－12　rRNA 操纵元的 RNA 转录子

在 16S 和 23SrRNA 的间隔序列中有 2 个 tRNA 分子；下图示 tRNA 的次级结构及其加工之前两侧的结构；箭头示加工位点，Ⅲ示 RNA 酶Ⅲ，P 为 RNA 酶 P

前体 rRNA 中的 tRNA 也必须通过切割才能形成成熟的 tRNA 分子。细菌细胞中还有其他一些 tRNA 基因作为一个转录单位进行转录，形成很大的前体 tRNA 分子；当然细胞中也有一些 tRNA 基因单独进行转录。所有这些前体 tRNA 转录子都必须经过切割加工，才形成成熟的 tRNA。前体 tRNA 的分割主要由 RNA 酶 P 和 RNA 酶 D 所催化，前者为一种内切核酸酶，在 tRNA 的 5′端切割前体 RNA，而后者则为一种外切核酸酶，水解从前体 RNA3′端到 tRNA3′－CCA 三核苷酸之间的磷酸二酯键。所有 tRNA 都具有－CCA 序列，它是 tRNA 分子连接氨酰基团的重要结构。各种 tRNA 从初级转录子上分割下来后，还必须经过进一步修饰才成为完全成熟的 tRNA，因为转录子中某些位点上原始的正常碱基 A、G、C 和 U 还需进一步修饰成某些稀有碱基（第一

章)。这些修饰过程均由不同的酶所催化。

三、翻译启始

(一) 翻译启始复合体的结构

细菌中大多数基因都转录成 mRNA，各种 mRNA 都具有一个核糖体结合位点，叫做 SD 序列（Shine－Dalegarno sequence)，其同感序列为 AGGAGGU。这段序列位于翻译起点密码 AUG 上游几个碱基的位置。核糖体结合位点的序列能够与 16SrRNA3′端的 7 个碱基 ACCUCCU 互补（图 3－13）使 mRNA 准确地结合到核糖体上。原核生物中，当转录机器以 DNA 为模板延长 mRNA 链时，翻译就已经开始。翻译启始复合体由 70S 核糖体、启始 tRNA（fmet－tRNA）和 mRNA 三部分组成，当 70S 复合体与下一个 tRNA 结合时，翻译过程进入肽链延长期。

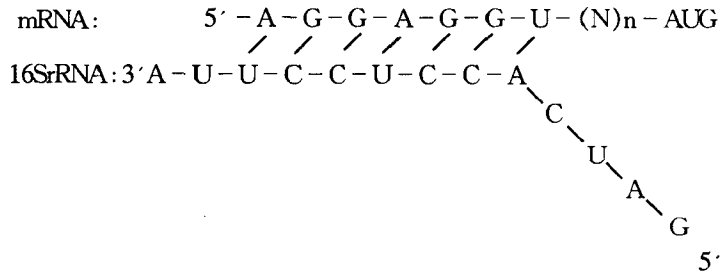


图 3－13　原核生物 mRNA 的 SD 序列与 16S rRNA3′端序列互作

虽然 30S 小亚基与翻译起始过程有关，但它本身并不能使 mRNA 与 tRNA 结合，翻译启始过程需要其他一些叫做启始因子（initiation factor，IF）的蛋白质。启始因子只存在于 30S 亚基上，当 30S 亚基与 50S 亚基结合，形成 70S 的核糖体后，它们又从小亚基上释放下来，因此它们不参与翻译延长过程（图 3－14)。

细菌中存在三种翻译启始因子，即 IF－1、IF－2 和 IF－3。IF3 是小亚基与 mRNA 中核糖体结合序列结合所必需的。IF－2 与启始 tRNA 结合有关，并控制其进入核糖体。IF－1 结合到 30S 小亚基上后，可能只对整个翻译启始复合体起着稳定作用。

根据对大肠杆菌 mRNA 核糖体结合位点的研究，翻译起点密码中 91％为 AUG，8％为 GUG，1％为 UUG。一般认为具有后 2 个翻译起点密码的基因只确定一些微量蛋白质的合成。例如，当细菌 lacZ 基因的起点密码 AUG 突变成 GUG 时，该基因的翻译效率降低 25％。

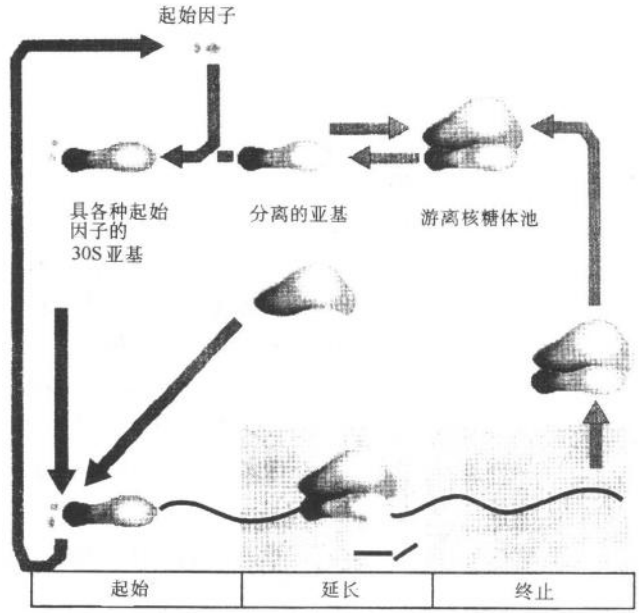


图 3－14　细菌翻译启始与核糖体亚基循环

起点密码 AUG 与 SD 序列之间的距离也与翻译水平的高低密切相关。例如在 T4 噬菌体 rⅡB 基因中，翻译启始复合体中 mRNA 正常的序列为 UAAG-GAaaauuAUG（其中小写字母为间隔区），如果 AUG 密码突变成 AUA，该基因则不能在大肠杆菌细胞中合成蛋白质。如果正常起点密码之前的一个碱基 U 突变成 G，则形成一个新的起点密码，这个 AUG 与 SD 序列之间只相隔 2 个核苷酸，其结果是这种 rⅡB 突变体所合成的蛋白质只有正常 rⅡB 蛋白的 10％。又如，编码 β－半乳糖苷酶的顺反子有 2 个 AUG 起点密码，二者距 SD 序列分别为 7 个碱基和 13 个碱基，其中相距较近的一个起点密码的使用次数是另一密码的 2～4 倍。根据对统计资料分析表明，SD 序列距起点密码 AUG 的最佳距离为 9 个碱基，如果二者之间的距离小于 5 个碱基，翻译过程将受到严重影响。

第二密码以及 mRNA 中其他核苷酸对调节翻译启始也可能具有某种影响。在许多高度表达的基因中，第二密码通常为 GCU，可能第二密码对 tRNA、核糖体等结合具有某种作用。另外，在 mRNA 的 －20 至 ＋13 之间的一段序列中，除了 SD 区的 G 序列、启点密码和第二密码 GCU 以外，其中多个碱基 A 对翻译效率也有重要影响。其作用可能是防止 mRNA5′序列与 16S rRNA 进行不准确配对。同真核细胞的 mRNA 一样（第四章）原核细胞 mRNA 在 －3 位

也几乎都是 A，而从 +10 至 +13 位则通常都是 UUAA，可能这些结构对 mRNA 与核糖体结合都具有某种促进作用。

（二）mRNA 的次级结构决定翻译起始位点的使用

大多数 mRNA 分子的核糖体结合位点的序列都能折叠成次级结构，如果这种次级结构不打开，核糖体就不能启始蛋白质合成。因此，这种次级结构的形成和熔解过程实际上就是一种特殊的翻译调节信号。例如细菌的一种抗药性质粒（R 质粒）有一抗红霉素（erythromycin）基因，该基因编码一种使 23S rRNA 上腺嘌呤基团甲基化的酶，阻止红霉素与核糖体结合，从而避免这种抗生素对蛋白质合成的抑制作用。在这种抗红霉素甲基化酶 mRNA 之前有一段编码前导多肽的前导序列，当核糖体翻译了前导多肽并脱离 RNA 之后，前导序列形成一种 RNA 发夹结构，掩盖甲基化酶的翻译启始序列，从而阻止该酶的翻译（图 3－15）。但是，当存在红霉素时，红霉素结合到核糖体上并延缓核糖体合成前导多肽的速度，此时就会形成另一种次级结构，使甲基化酶的翻译启始区暴露在外，从而进一步合成甲基化酶。如果甲基化酶的翻译水平始终很低，那么合成的抗红霉素核糖体也就很少，甲基化酶顺反子就能继续翻译，抗红霉素的核糖体也就愈多。当抗红霉素的核糖体数目达到一定程度时，这种核糖体翻译前导肽链，mRNA 上不再含有滞留的核糖体，于是最后关闭甲基化酶的合成。因此这种甲基化酶合成的诱导过程是一种自动调节的过程。

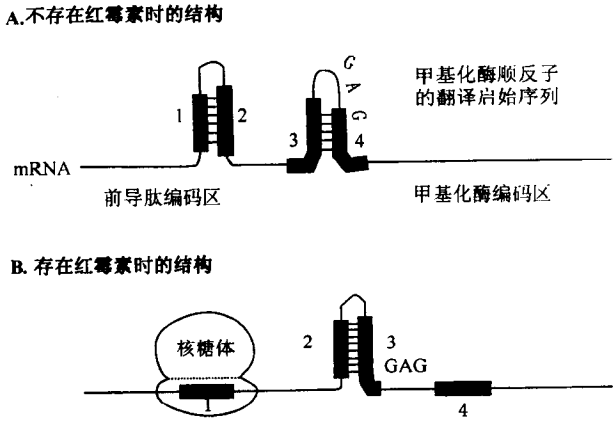


图 3－15　抗红霉素甲基化酶基因表达的自动调节

A. 核糖体翻译前导肽至前导序列的末端，脱离 RNA；此时形成 1、2 发夹和 3、4 发夹，从而阻止核糖体附着到甲基化酶启始序列上，不能翻译甲基化酶顺反子　B. 核糖体在翻译前导多肽的过程中被红霉素阻隔，从而遮盖序列 1，形成 2、3 发夹，释放甲基化酶和翻译启始序列 4，并使之与第二个核糖体结合，翻译甲基化酶

mRNA 的次级结构调节翻译过程的另一例子见于 T4 噬菌体的溶菌酶基因的翻译。T4 噬菌体的溶菌酶基因在侵染的早期和后期都能转录，但是其早期 mRNA 完全不能翻译。进一步研究知道，早期转录子能形成一种稳定的次级结构，其中包含了 SD 序列和起点密码 AUG，而侵染后期转录的 mRNA 则不含这种具有抑制作用的次级结构，故能正常翻译。

（三）多顺反子 mRNA 的不等翻译及翻译调节

虽然原核生物的 mRNA 通常都是多顺反子结构，但是，由同一条 mRNA 分子编码的各种蛋白质并不一定等量合成。根据对乳糖操纵元的研究（见后），β－半乳糖苷酶的合成量比半乳糖苷渗透酶和半乳糖苷转乙酰酶的合成量高得多，这三种蛋白质的比例分别为 1∶1/2∶1/5。对于这种现象有三种解释，第一种解释是核糖体同 RNA 分子上不同顺反子的翻译起点结合的速率不同，其结合过程依赖于 AUG 起点密码之前重要的结合序列。第二种可能性是，核糖体只与 β－半乳糖苷酶顺反子之前的位点结合，其下游顺反子的翻译取决于在阅读完上游顺反子的终止密码之后核糖体脱离 mRNA 的速率。第三种解释是影响翻译的因素是细胞中与某些密码子相对应的 tRNA 的含量很低，造成核糖体在这些密码上停留的时间较长，以待某些稀有 tRNA 扩散进来，故翻译效率较低。

另一方面，在多顺反子 mRNA 的翻译过程中，有时 3′顺反子往往比 5′顺反子具有更高的翻译效率。这种现象说明，多顺反子 mRNA 中可能存在一些潜在的翻译启始位点。例如，当大肠杆菌的 lacI 基因中发生无义突变后，核糖体可以跳过这一终止密码而再度启始翻译，这种位点叫做再启始位点（reinitiation site）。再启始位点不含 SD 序列和起点密码，正常情况下这些位点并不活动，当翻译过程在其附近中止后，使核糖体浓度局部升高而形成启始复合体，此时这些位点才具有活性。这种现象叫做翻译偶联（Translationcoupling）。例如在 trp 操纵元中（本章第三节），trpD 与 trpE 偶联，trpA 与 trpB 偶联。

两个编码序列之间的翻译偶联现象已用于通过大肠杆菌来大量生产小牛生长激素（bovin growth hormon BGH）。在构建融合基因（chimeric gene）的过程中，直接将 BGH 编码序列置于 SD 序列下游并不能使该基因高水平地翻译，但是，如果在 AUG 和 BGH 编码序列之间再加上一段额外核苷酸序列，bgh 基因则能高水平地翻译，甚至在 bgh 基因之前有终止密码时，该基因也能高水平表达。

某些蛋白质能够抑制某种特殊 mRNA、多顺反子 mRNA 或某一组无关联的 mRNA 的翻译过程。这些蛋白质能够识别目标 mRNA 的起点密码周围的某种短的特殊序列，这种序列有时叫做翻译操作子（translational operator）。例如，噬菌体 MS2 的复制酶 mRNA 上有一段能形成发夹结构的序列，外壳蛋白

能识别这一位点的序列，抑制复制酶 mRNA 的翻译。因为这类抑制物直接阻碍核糖体通过目标结合位点。

噬菌体 T4 的 regA 基因编码的蛋白质为一种翻译抑制物，它能同许多 T4 的 mRNA 结合。例如该蛋白在 rⅡBmRNA 的结合位点的序列与起点密码重叠，其结构为 UAUGU。regA 蛋白还有另一种功能，即活化某些特殊 mRNA 的翻译。例如将载有 regA 基因的质粒引入大肠杆菌后，细菌中有些蛋白质的合成受到抑制，而另一些 mRNA 的翻译则受到活化。

翻译过程还可受某些 RNA 分子的调节，例如转座子 Tn10 的 mRNA 在翻译过程中可受一种与其互补的 RNA 的抑制。这种通过 RNA－RNA 互补（结合）来抑制翻译过程可能是关闭基因表达的方式之一。

（四）反义 RNA 对翻译的调节

大多数反式作用调节子都是蛋白质，但是某些小 RNA 分子也可以调节基因表达。同蛋白质调节子一样，RNA 也是独立合成的分子，它通过扩散，到达其靶位点。调节子 RNA 的靶位点为一种单链核苷酸序列，调节子通过与靶位点的序列互补，形成双链区而行使功能。这种双链区有两种作用，其一是通过形成特殊位点，阻止靶位点的序列行使功能。其二是靶位点所在区域的一部分序列形成双链后会改变另一区域的构型，从而间接地影响其功能。所以调节子 RNA 的总体作用是使靶序列形成次级结构来控制其活性。

反义 RNA（antisense RNA）就是细菌细胞中的 RNA 调节子，其调节方式有三种（图 3－16）。第一，反义 RNA 与 mRNA 结合，堵塞核糖体结合位点，从而阻止蛋白质合成的启始。第二，反义 RNA 通过与 mRNA 结合，形成内切核酸酶作用的位点，直接破坏 mRNA 的稳定性。第三，反义 RNA 可能通过与 RNA 结合，形成类似终止子的结构，引起转录提前终止，从而影响基因表达。

根据对大肠杆菌的控制细胞对渗透压作出反应的一组基因的研究，一种小分子的反义 RNA 直接控制这组基因的表达（图 3－17）。大肠杆菌有 2 个外膜蛋白基因 ompF 和 ompC 的表达受培养基渗透性调节，另一基因 envZ 编码一种接受渗透压信号的受体分子。当渗透压增加时，EnvZ 蛋白活化由 ompR 基因编码的蛋白质。OmpR 蛋白为一种正调节子，活化结构基因 ompC 和另一调节基因 micF 的转录。micF 的编码产物为 174nt 的 RNA，它是一种反义 RNA，即一种能与 mRNA 进行碱基互补配对、阻止 mRNA 翻译的 RNA。所以当渗透压增加时，就会促进 MicF 反义 RNA 的合成，反义 RNA 然后关闭 ompF 的 mRNA 的翻译。MicF RNA 与 OmpF mRNA 中包括核糖体结合位点在内的一段序列结合，阻止 ompF mRNA 翻译。还有一种可能性是，当 ompF mRNA 形成双链区以后，其稳定性受到破坏，易于遭受切割双链 RNA 的核酸酶的切割。

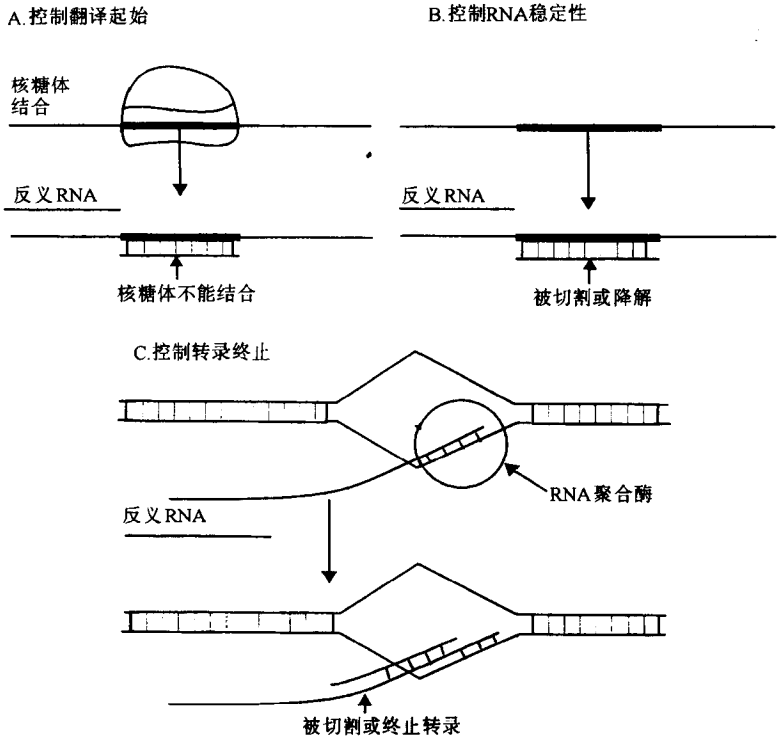


图 3－16　反义 RNA 调节基因表达的方式

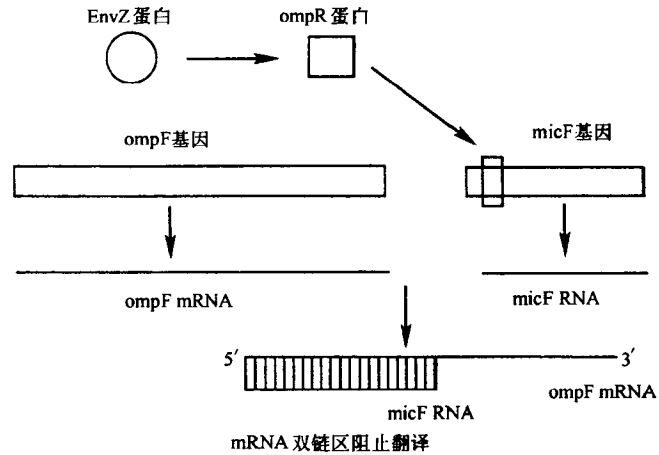


图 3－17　反义 RNA 调节细菌 mRNA 翻译的过程

渗透压升高，活化 EnvZ 蛋白，然后 EnvZ 蛋白活化 OmpR 蛋白，OmpR 活化 micF 基因转录，micF RNA 与 OmpF mRNA 的 5′区域结合，阻止其翻译

第三节　操纵元的调节表达

在细菌中，绝大多数编码蛋白质的基因都以操纵元的形式组织在一起，许多操纵元在调节表达方面都有相似之处，但不同操纵元也都有各自不同的结构特点和调节表达方式。

一、操纵元调节表达的一般模型

如第二章所述，根据操纵元在一般情况下是否表达，我们将操纵元分成诱导型操纵元和抑制型操纵元。在诱导型操纵元中，有些操纵元的操作子通常与一种抑制蛋白结合，所以在正常情况下并不表达，这类操纵元叫做负控制诱导型操纵元。只有当某种诱导物（一种小分子化合物）与抑制蛋白（反作用子）结合后，抑制物的构型发生变化并脱离操作子，操纵元才能转录。

还有一类诱导型操纵元在正常情况下也不表达，当一种小分子的辅活化物与一种活化蛋白反作用子结合后，使活化蛋白的构型改变并结合到操纵元 5′ 端附近某个位点上，操纵元才能转录，这类操纵元叫做正控制诱导型操纵元。

抑制型操纵元只有一种，即负控制抑制型操纵元。这种操纵元一般都能转录，当一种小分子的辅抑制物与一种抑制物蛋白结合后，这种复合体再结合到操作子上而抑制转录过程。各类操纵元表达的调节因子见表 3－4。

表 3－4　部分操纵元的调节因子

操纵元	诱导物	辅活化物	辅抑制物	操纵元类型
lac	异丙基硫代半乳糖苷别乳糖	cAMP	—	负诱导或正诱导
gal	半乳糖	cAMP	—	正诱导或负诱导
mal	麦芽糖	cAMP	—	正诱导
hut	尿利酸盐	cAMP	—	负诱导或正诱导
trp	—	—	色氨酸	抑制
arg	—	—	精氨酸	抑制

操作子通常位于启动子至转录起点之间，一般认为抑制物与操作子结合后形成一种物理障碍，阻止 RNA 聚合酶通过启动子，到达转录起点。抑制物由

结构基因编码，这些基因可能位于其控制的操纵元附近，也可能距离所控制的操纵元很远，每一细胞中通常都只有少数几个拷贝。

还有一些操纵元的调节表达方式与上述操纵元有所不同，如编码与某些氨基酸合成和其他一些基本代谢产物合成的酶类的操纵元等，这类操纵元通过叫弱化反应（attenuation）的调节方式进行表达（将在本节以后讨论）。

一般说来，抑制型操纵元编码一些与小分子化合物合成有关的酶类，如氨基酸、嘌呤和嘧啶碱合成有关的酶。这些小分子化合物是细胞合成代谢所必需的，在细胞所生活的环境中都可能不同程度地存在着这类化合物。如果这类化合物在周围环境中含量充足，与这些化合物合成有关的酶类的操纵元就停止表达。相反，诱导型操纵元编码一些与分解代谢有关的酶类，如分解细胞中某些过量存在的糖类和氨基酸等的酶类。如果缺少这些基质，诱导型操纵元则不表达。

根据对操纵元中各个调节区突变的研究，编码抑制物、辅抑制物和活化物的基因发生突变后，依突变位点不同，就会对操纵元表达产生各种不同影响，这些突变主要影响其编码产物与操作子互作。因此，抑制物改变后就不再与操作子结合，在这种情况下，负控制诱导型操纵元就表现为组成型表达。相反，如果某个突变不能使诱导物与抑制物结合，操纵元将永远不能表达。同样，操作子发生突变也会明显影响抑制物以及其他一些调节蛋白的结合过程。这些突变分别见表 3－5 和表 3－6。

表 3－5　编码调节蛋白基因发生突变（反作用子突变）的效应

调节蛋白	突变的抑制功能	结　果
抑制物	与 DNA 结合	操纵元组成型表达
抑制物	与诱导物结合	不能诱导操纵元表达
活化物	与 DNA 结合	不能诱导操纵元表达
活化物	与辅活化物结合	不能诱导操纵元表达
活化物	与 DNA 结合	抑制操纵元表达

表 3－6　操作子突变（顺作用子突变）的效应

操纵元类型	可能的突变效应	结　果
负控制诱导	控制抑制物的结合	操纵元组成型表达
正控制诱导	控制活化物与辅活化物结合	操纵元不表达
负控制抑制	控制抑制物与辅抑制物结合	对操纵元失去抑制功能

所有正控制诱导型操纵元都有一个共同特征，即诱导作用是通过叫做分解代谢活化蛋白（catabolite activator protein，CAP）或叫做环化 AMP 受体蛋白的蛋白质与 DNA 上调节位点结合而实现的。CAP 蛋白有一个 cAMP 结合位

点，当与 cAMP 结合后，只与 DNA 上调节位点互作。这些 CAP 蛋白－cAMP 结合位点通常位于 RNA 聚合酶识别位点（－35 框）上游，并与之紧邻的位置。一般认为，CAP 蛋白复合体与这一位点结合有利于 RNA 聚合酶结合进来。有些 CAP 蛋白结合位点的序列具有一段长约 10bp 的对称结构（图 3－18），这种结构对 CAP 蛋白识别适当的 DNA 序列具有重要作用。如果编码 CAP 蛋白的基因或编码腺苷环化酶（adenylate cyclase，CYC）的基因发生缺陷型突变，操纵元就不能表达。

当细菌生长的环境中葡萄糖的浓度很低时，腺苷环化酶受到刺激，催化合成 cAMP，然后 cAMP 再结合到 CAP 蛋白上，最后 CAP 蛋白－cAMP 复合体再结合到诱导型操纵元的相应位点上，使操纵元指导合成一些与能源代谢有关的酶类，如乳糖、半乳糖、阿拉伯糖和组氨酸等分解代谢有关的酶类。

GTGAGTTAGCTCAC
CACTCAATCGAGTG　　乳糖操纵元

GTGGAAACGATTCCAC
CACATTTGCTAAGGTG　半乳糖操纵元

图 3－18　乳糖和半乳糖操纵元中 CAP 蛋白复合体结合位点的颠倒对称结构
碱基对中间的线条示对称的碱基序列

二、乳糖操纵元的负控制和正控制诱导表达

乳糖操纵元（lac operon）含有三个结构基因，它们分别编码与乳糖的吸收和分解有关的酶类。在这三个结构基因的上游依次为操作子、启动子、CAP 蛋白结合位点以及一个编码抑制蛋白的结构基因即所谓的调节基因（图 3－19）。由 lac 操纵元编码的酶的合成既可以被诱导也可以被抑制。如果将大肠杆菌细胞在只以葡萄糖作为碳源和能源的培养基中培养，其细胞中 β－半乳糖苷酶和乳糖渗透酶的活性都很低。如果将这些细胞再转移到含有乳糖的培养基上，则 β－半乳糖苷酶和乳糖渗透酶的活性在 20 分钟内可增加一千倍。早期研究表明，β－半乳糖苷酶活性增加是由于合成了新的 β－半乳糖苷酶分子，在将细菌从葡萄糖培养基转移到乳糖培养基几分钟内即开始合成。如果将细菌从含有乳糖的培养基转移到含有葡萄糖的培养基上，β－半乳糖苷酶的合成则迅速被关闭或被抑制。当细胞在以乳糖为唯一碳源的培养基上生长时，乳糖渗透酶和硫代半乳糖苷转乙酰酶也很快被诱导合成，当细胞被转移到没有乳糖的培养基上时，这两种酶的合成又迅速被抑制。

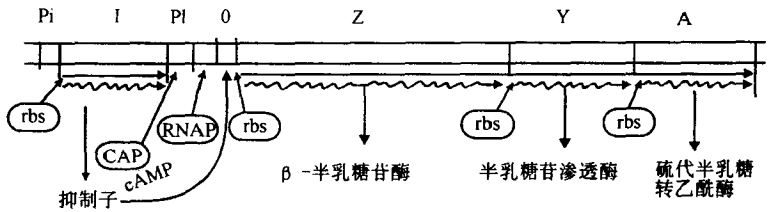


图 3－19　lac 操纵元

RNA 聚合酶（RNAP）在启动子 PI 位点启始 lacZ、Y 和 A 基因转录。分解代谢基因的编码产物 CAP 与 cAMP 形成复合体，对这一过程进行正调节；lacI 基因的产物（抑制物）对这一过程进行负调节，抑制物反过来又受诱导物的抑制。核糖体（rbs）结合到每一基因的翻译启始信号上，启始各种蛋白质的翻译（波浪线箭头）

利用一种无色化合物 5－溴－4－氯－3－吲哚－β－D－半乳糖苷(5－bromo－4－chloro－3－indoyl－β－D－galactose,通常叫做 x－gal)很容易测定 β－半乳糖苷酶和乳糖渗透酶顺反子的突变。x－gal 可由 β－半乳糖苷酶水解成一种不可溶的蓝色化合物而又不影响细胞生长,因此通过颜色就可测定到细胞中的 β－半乳糖苷酶。例如当培养基中加入 x－gal 以后,乳糖渗透酶将其转运到细胞内,然后由 β－半乳糖苷酶将其水解成蓝色化合物。所以,能够产生 β－半乳糖苷酶的细胞和由其形成的菌落成为蓝色,而不能合成 β－半乳糖苷酶的菌落则为白色,二者很容易区别。根据这种原理,现在已研究清楚了调节乳糖操纵元表达的反作用蛋白和顺作用 DNA 序列。lac 操纵元控制表达有两种方式,第一种方式是当培养基中存在乳糖时,抑制蛋白脱离操作子,这就是所谓的抑制蛋白的负控制诱导;第二种方式是由 CAP 蛋白与 DNA 结合所调节的正控制诱导。

在负控制诱导表达方式中，当 lacI 基因编码的抑制物结合到操作子上以后，阻止 RNA 聚合酶与启动子（Pi）形成开放复合体；当诱导物与抑制物结合后，引起抑制物的构型异构化，使之与操作子亲和性降低，脱离操作子，RNA 聚合酶开始转录下游结构基因。

lac 操纵元中，同抑制蛋白结合的操作子位点大多数都位于起点密码至转录起点之间，其中含有一个颠倒对称结构，这种结构可能对促进抑制物结合具有重要作用。操作子与启动子部分重叠，这种位置特征也与结合机制一致，即当抑制物结合进来后，RNA 聚合酶就不能结合。

抑制物基因的终止密码位于近 CAP 结合位点的 5′端，由于该基因之前有一个很弱的启动子 Pi，因此其转录水平很低。据分析，在一个大肠杆菌细胞中，大约只有 10 条抑制蛋白分子，但是由于它与 lac 操纵元的操作子具有很高亲和性，因此在缺少乳糖的条件下，就足以能够抑制 lac 操纵元表达。

lac 操纵元的实际诱导物并非乳糖，而是一种叫做别乳糖（allo－lactose）

的异构体，由乳糖合成而来，其合成过程由操纵元编码的第一种酶即β－半乳糖苷酶（β－galactosidase）所催化，同乳糖一样，别乳糖也可以由β－半乳糖苷酶水解。此外，还有其他一些人工合成的化合物如异丙基－β－硫代半乳糖苷（isopropyl－β－thiogalactoside，IPTG）也能诱导 lac 操纵元表达，这些诱导物统称为安慰诱导物或义务诱导物（gratuitousinducer）。IPTG 是一种对乳糖操纵元进行遗传研究的非常有用的诱导物，它可以扩散到细胞内，而且不被降解，所以在试验期间其浓度始终保持不变。

大肠杆菌中有关操纵元的正控制诱导表达机制目前还不很清楚，正活化物 CAP－cAMP 复合体在调节 lac 操纵元表达过程中起着三种可能的作用。第一，这种正活化物可能通过直接与 RNA 聚合酶互作，提高 RNA 聚合酶与启动子的亲和性和促进关闭复合体向开放转录启始复合体的转变即转录活化的补给模型。第二，正活化物通过扰乱启动子的 DNA 构型而调节转录，因为在各种操纵元中，CAP 结合位点的位置各不相同。第三，CAP－cAMP 复合体行使一种抑制物的功能，可能抑制 RNA 聚合酶与某个位点的无效结合，这一位点与启动子竞争 RNA 聚合酶，可能 lac 操纵元的 P_2 启动子就是这种位点（图 3－8）。P_2 与启动子 P_1 重叠，只在离体条件下启始 lac 操纵元的另一转录子的合成，但在活体内不能启动转录过程。

三、色氨酸操纵元与弱化反应

细菌基因组中存在数种与氨基酸合成有关的操纵元，其转录既可以受氨基酸自身的抑制，也可以受有关氨基酸衍生物的抑制，所以只要培养基中有足量氨基酸存在，这些操纵元就无须表达。

色氨酸操纵元（trp operon）是其中之一，它含有 5 个结构基因，分别编码邻氨基苯甲酸合成酶Ⅰ（anthranilate synthetase）和合成酶Ⅱ、磷酸核糖邻氨基苯甲酸异构酶（phosphoribosyl－anthranilate isomerase）以及色氨酸合成酶（tryptophan synthe－tase）Ⅰ和Ⅱ（图 3－20）。这些酶参与由分支酸盐（chorismate）合成色氨酸的各个步骤。编码抑制物的基因（trpR）不与 trp 操纵元相连，当抑制物与色氨酸结合后，形成的复合体然后再与邻近转录起点 5′端的操作子结合，阻止 RNA 聚合酶通过启动子，关闭转录过程。因此，色氨酸本身并不是一种诱导物，而是一种辅抑制物，即它活化一种特殊的抑制物，从而阻止色氨酸操纵元的转录。

Trp 操纵元的转录起点与其结构基因的翻译起点密码之间相距 162bp，这段前导序列对 trp 操纵元的第二种控制方式极为重要。RNA 的前导序列含有一

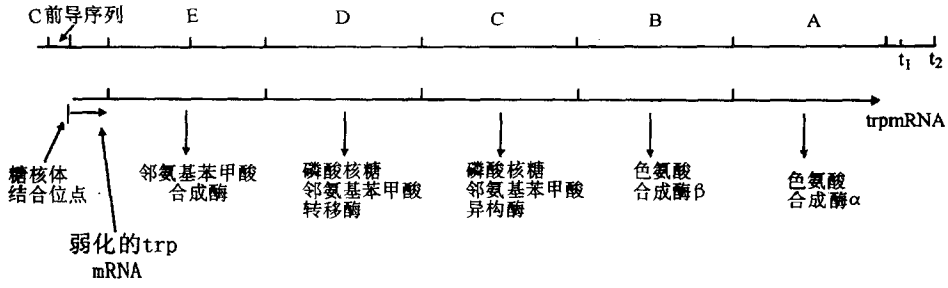


图 3－20　色氨酸操纵元的组织

示 5 个结构基因和两个终止子（t_1 和 t_2），其中 t_1 不需要 ρ 因子，而 t_2 则只有存在 ρ 因子时才具有功能

个长 42bp 的可读框，其中存在两个相邻的色氨酸密码，可读框的 3′方向存在有一个典型的转录终止信号，即一段富含 G－C 碱基对的区域，这段序列可以形成发夹结构，其后为一段 U 碱基串。这种转录子可以通过分子内碱基配对形成两种不同的次级结构（图 3－21），由于这两个发夹结构共用主杆的一部，所以二者不能同时形成。在活体条件下，新合成的 RNA 与核糖体结合并合成由前导序列编码的一段短的肽链。虽然 RNA 上存在核糖体，但仍然能够形成结构Ⅰ中终止子的主杆和环而终止转录过程。当细胞中缺少色氨酸和负载的色氨酸 tRNA 时，核糖体将被阻隔在两个色氨酸密码之一上，形成一种新结构（结构Ⅱ）即形成反终止子的主杆和环，而不能形成终止子结构，因此转录过程继续进行。

这种类型的控制依赖于细胞中某种特殊的氨酰 tRNA（aminoacyl－tRNA）的浓度，也就是说当色氨酸含量丰富，载有色氨酸的氨酰 tRNA 浓度高时，由前导序列中可读框编码的 RNA 能够正常翻译，从而抑制 trp 操纵元的其他部分转录，所以就把这种现象叫做弱化作用（attenuation），而把这段能中止 RNA 转录的区域叫做弱化子或弱化基因（attenuator）。据分析，抑制物/色氨酸复合体与操作子结合可使 trp 操纵元的转录降低 70 倍，而弱化作用则可使该操纵元的转录降低 80 倍。

格兰氏阳性细菌枯草杆菌（*Bacillus subtilis*）的色氨酸操纵元（trp EDCFBA）的表达也受转录弱化作用调节。但不同的是，前导区（长 204 个碱基）中终止子和反终止子的形成需要前导区 RNA 与一种受色氨酸活化的 RNA 结合蛋白（TRAP）互作。当培养基中存在色氨酸时，TRAP 与前导区 RNA 结合，诱导形成转录终止子，操纵元不继续转录。当不存在色氨酸时，TRAP 就不与前导区结合，从而形成反终止子（图 3－22），该操纵元转录。枯草杆菌的 trp 弱化作用无需前导 RNA 进行翻译，而大肠杆菌的 trp 弱化机制则是通过核糖体翻译出一段前导肽来控制弱化反应。

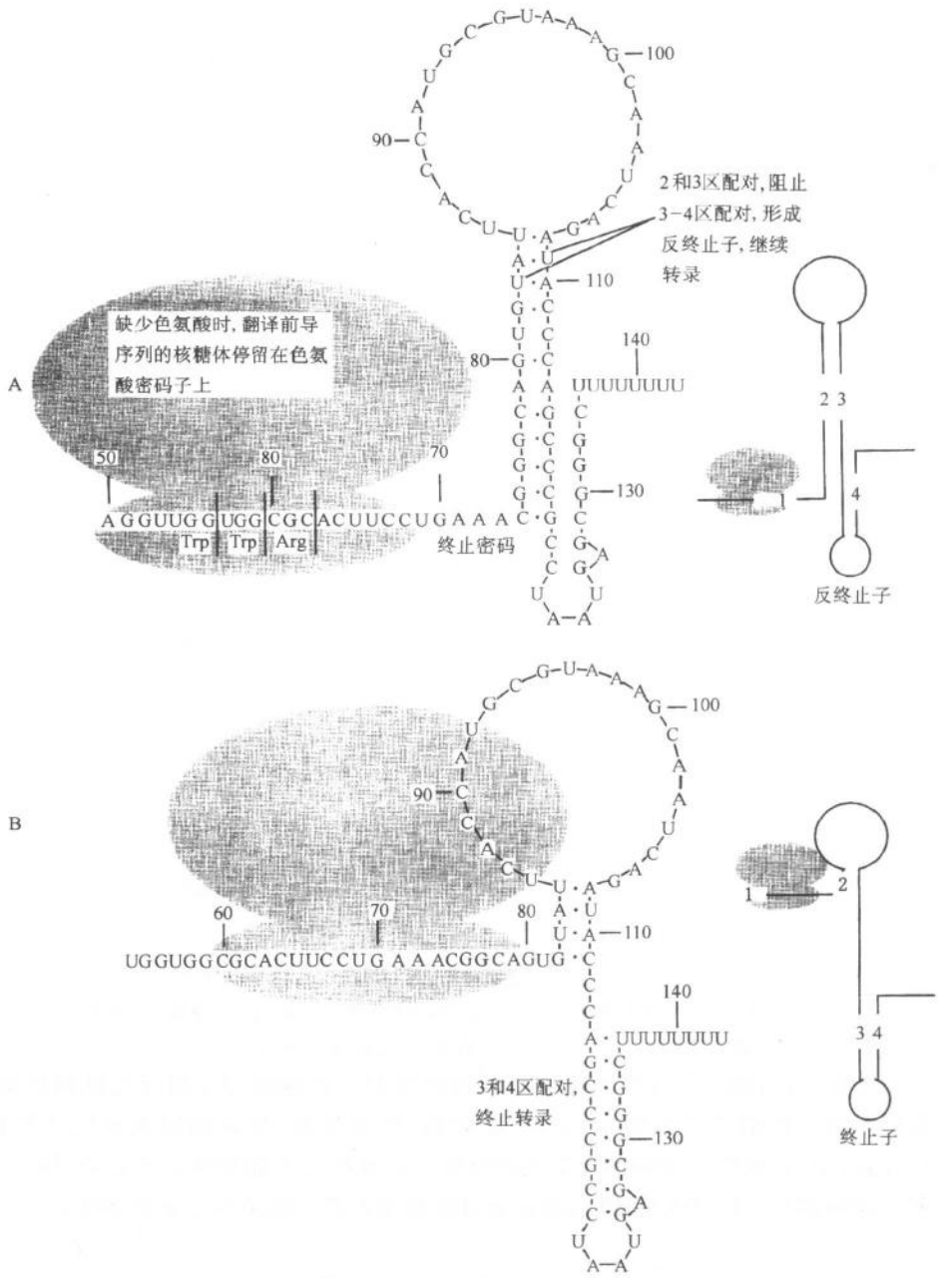


图 3－21　大肠杆菌色氨酸操纵元的弱化作用模型

A. 缺少色氨酸时的反终止作用　B. 存在色氨酸形成终止子

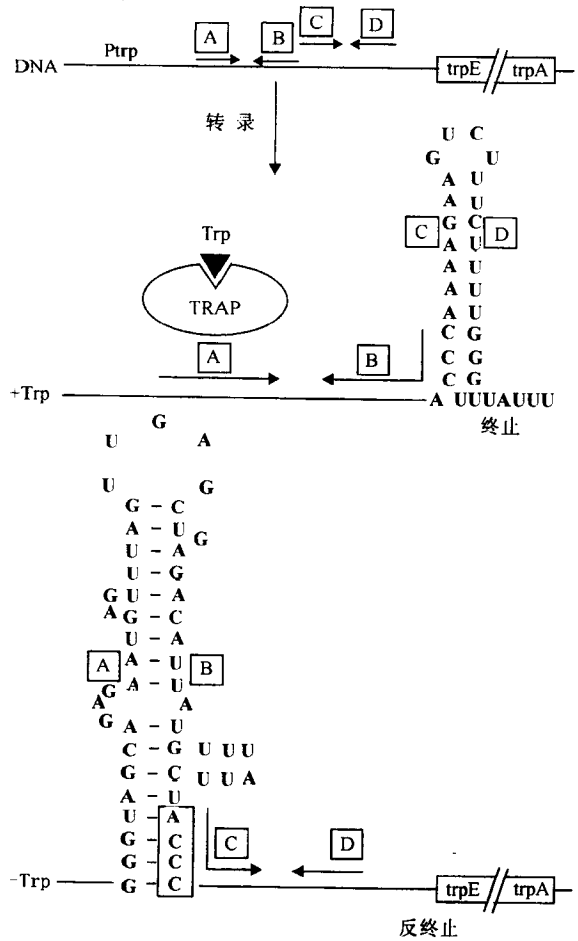


图 3－22　枯草杆菌色氨酸操纵元调节表达的转录弱化模型

RNA 结合蛋白须与色氨酸结合，然后这种复合体再与 A 区结合，诱导终止子形成（方框内的字母示终止子和反终止子的互补链，根据 G.Gollnick，1994）

现在知道，细菌中许多其他氨基酸的合成都与色氨酸的控制合成机制相似，包括大肠杆菌的苯丙氨酸、苏氨酸、亮氨酸、异亮氨酸、缬氨酸以及沙门氏杆菌（*Salmonella typhii*）的组氨酸和亮氨酸合成。在这些氨基酸操纵元的转录子中，其前导序列都具一段短的可读框，都能通过碱基配对形成转录终止子和反终止子。

四、核糖体蛋白操纵元与翻译自动调节

在大肠杆菌中，70S 核糖体由 55 种不同蛋白质组成，其中大亚基有 31

种，小亚基有20种，另外几种为大小亚基所共有。编码这些蛋白质的基因连锁在几个不同的操纵元之中（表3-7）。其中有些操纵元还编码另一些与转录和翻译过程有关的蛋白质。

表3-7　大肠杆菌核糖体蛋白操纵元

操纵元名称	基因及其产物	调节蛋白
L11	rplK rplA rplJ rplL rpoB rpoC L11 [L1] [L10] L7 β β⁻	L1 L10（调节 rpl_J 下游的几个顺反子）
Str	rpsL rpsG fusA tufA S12 [S7] EF-G EF-Tu	S7
S10	rpsJ rplC rplB rplD rplW rplS rplV rpsC rplP rpmC rpsQ S10 L3 [L2] [L4] L23 S19 L22 S3 L16 L29 S17	L4
SPC	rplN rplX rplE rpsN rpsH rplF rplR rpsE rpmD rplO L14 L24 L5 S14 [S8] L6 L18 S5 L30 L15	S8
S16	rpsP orf trmD rplQ S16 L19	不清
α	rpsM rpsK rpsD rpoA rplQ S13 S11 [S4] α L17	S4

注：线上为基因符号，线下为蛋白质，L和S分别为大亚基和小亚基，EF为蛋白质合成延长因子，加框的蛋白质可抑制所在操纵元的转录，trmD编码某些tRNA中碱基G的甲基化酶。

现在知道，大肠杆菌中所有核糖体蛋白都是等量合成，但其合成的调节机

制目前还不很清楚。其中有些核糖体蛋白操纵元的转录受到由自身结构基因编码的一种蛋白质的抑制，这种蛋白质与该操纵元的5′侧翼区结合而抑制转录过程。这种机制可以保证细胞不致于合成过量核糖体蛋白，当细胞中核糖体蛋白达到一定浓度时，其合成过程自动关闭。

当细菌处于正常生长条件下时，在L11操纵元中，4种核糖体蛋白的合成量比RNA聚合酶亚基的合成量大得多。根据DNA序列分析知道，在rpl和rpoB基因之间存在一大段非翻译的间隔序列，这段序列含有一个相当弱的终止子（或弱化子）结构，它可以在转录进行到RNA聚合酶基因之前终止翻译/转录过程。并且还有一个与23S和16SrRNA基因间序列中主杆和环状结构非常相似的结构，它可能是RNA酶Ⅲ切割初级转录子的位点，通过切割，将RNA聚合酶的mRNA与核糖体蛋白的mRNA分开，使之分别进行翻译。另外，这一区域含有两个普里布诺框，它们可以独立地作为RNA聚合酶基因转录的启动子，因此L11操纵元实际上由2个独立的操纵元组成。

在编码核糖体蛋白S16和L19的S16操纵元中，这两种蛋白质的结构基因被编码转甲基酶（transmethylase）的基因和一个编码未知产物的可读框所隔开，其中转甲基酶的功能是使某些tRNA的鸟嘌呤基因甲基化，根据分析，每个细胞中两种核糖体蛋白大约各有8000个分子，而转甲基酶大约只有80个分子，究竟同一操纵元中两类不同基因在表达水平方面存在差异的原因是什么，目前还不清楚。转甲基酶的分子量比这两种核糖体蛋白大两倍多，所以其mRNA翻译持续的时间要长得多。另外，转甲基酶基因还使用几种其他高水平表达的基因不经常使用的密码。但是这两种核糖体蛋白单独是不可能使这一操纵元中转甲基酶基因的表达速率降低100倍的，可能在翻译水平上还存在其它一些控制机制。

表3－7中第二个操纵元str除了编码两种核糖体蛋白以外，还编码两种用于翻译过程的延长因子，其中EF－Tu的合成速率大约是EF－G和两种核糖体蛋白合成速率的7倍。由于tufA基因在距fusA基因3′末端50bp的位置有一个启动子，所以这一基因可以单独转录。在大肠杆菌染色体的另一位点，还有一个tufB基因，由这一基因编码的蛋白质与EF－Tu非常相似，在细菌细胞中其含量相当于EF－Tu的四分之一到二分之一。tufB基因所在的操纵元还编码4种tRNA，该基因位于操纵元3′端，这一操纵元是目前已知的惟一编码几种RNA和mRNA的操纵元。在这一操纵元中，在第一个tRNA基因的上游有一个很强的启动子，在最后一个tRNA基因和tufB基因之间还存在一个较弱的启动子，可能tufB基因能够独立于这一操纵元而单独转录。

如前所述，大肠杆菌的核糖体由55种蛋白质和3种rRNA组成，所有这

些成分都不是来自同一操纵元，编码 rRNA 的基因和核糖体蛋白的基因分散在细菌染色体的不同区域。如果所有这些操纵元或基因都协调表达，就绝不会出现过量核糖体蛋白或过量 rRNA，当其中一种成分合成时，其他各种成分也同时合成。协调蛋白质和 rRNA 合成的机制是，在核糖体形成过程中，某些核糖体蛋白结合到 rRNA 的某些特定区域，当 rRNA 上的结合位点全部被占用后，游离的核糖体蛋白就结合到自身的 mRNA5′端的有关位点上，阻止其 mRNA 进一步翻译。因此，这些核糖体蛋白又调节其自身 mRNA 的翻译（图 3－23）。

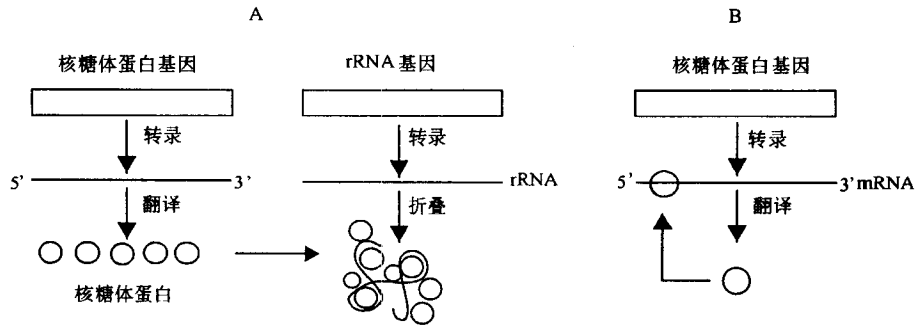


图 3－23　核糖体操纵元的翻译自动调节及其对 rRNA 水平的反应

A. 当存在 rRNA 时，核糖体蛋白就与其结合。当不存在游离核糖体蛋白时，核糖体 mRNA 继续翻译　B. 当不存在 rRNA 时，核糖体蛋白累积，并结合到自身 mRNA5′端，阻止其翻译

五、转录的总体控制

我们知道，控制某些特殊的分解代谢途径和合成代谢途径的基因的表达都是受一定的特殊机制控制的。不仅如此，细菌还能够从总体水平上对某些主要的环境变化作出反应。从进化的观点看，这对于细菌抵抗不良环境条件是有利的。例如，将在营养丰富的完全培养基中培养的细胞转移到营养匮乏的培养基中培养，细菌细胞对这种环境压就会作出反应，即所谓的严紧反应（stringent response）。严紧反应是指细菌在缺少氨基酸等营养的条件下，其细胞内发生的一系列协调反应过程，包括停止合成 rRNA、tRNA 和核糖体蛋白以及抑制碳水化合物、脂肪和肽聚糖（peptidoglycan）的合成，并同时显著提高某些氨基酸合成酶基因的转录水平的现象（表 3－8）。这一反应受一对四磷酸鸟嘌呤（ppGpp，3′－二磷酸－5′二磷酸鸟嘌呤）和五磷酸鸟嘌呤（pppGpp）所调节。

不表现出严谨反应的细菌突变体叫做松弛型菌株（relaxed strain），其突变位点发生在 relA 基因之中。relA 基因编码调节严紧反应的酶，即一种焦磷酸转移酶（pyrophosphoryl transferase），它与核糖体结合，在存在未负载氨基酸的 tRNA 条件下具有活性。未负载的 tRNA 通过与核糖体的 A 位点结合促进这种酶的合成，从而使 ppGpp 的浓度急剧增加 10 倍以上。但是有关 ppGpp 影响细菌基因组中某些操纵元转录的详细情况目前还不清楚。有些受严紧反应抑制的基因在转录起点附近都有一段结构保守序列 CGCCNCC，这段序列可能是 ppGpp 的结合位点，也可能是某种抑制蛋白的结合位点。另一方面，ppGpp 也可能与 RNA 聚合酶结合。一般认为 RNA 聚合酶可能采取几种不同构型以识别不同启动子，如果确实如此，ppGpp 的结合就可能稳定某些种类的 RNA 聚合酶构型，从而改变其转录某些基因或操纵元的效率。

表 3－8　严紧反应的某些效应

基　　因	对转录的影响
rRNA	受抑制
tRNA	受抑制
核糖体蛋白	受抑制
EF－Tu 和 EF－G	增强
lac 操纵元	增强
trp 操纵元	增强
半乳糖激酶	增强
核酮糖激酶	增强
氯霉素乙酰基转移酶	增强

对氨基酸饥饿信号 ppGpp 的发现表明，细胞中其他一些特殊代谢产物可能也具有相似作用，它们可能促进细胞中某些酶或蛋白质的合成，这类信号分子统称为报警激素（alarmone）。例如当细胞缺少 tRNA 时，氨酰 tRNA 合成酶就合成一种二腺嘌呤四磷酸（AppppA），并使其累积于细胞之中，给细胞提出警报。在真核细胞中，AppppA 刺激 α－DNA 聚合酶的活性，通常认为它是 DNA 复制受阻的一种信号，从而促进细胞进行新的 DNA 合成。

六、热激蛋白的合成

抑制子和具操纵元特异性的活化子通常都是通过改变启动子的启始效率和阻隔启动子来控制某些基因表达，这些抑制子和活化子并不改变 RNA 聚合酶的活性，但是，当环境条件发生急剧变化时，有些基因的表达就会采取另一种

完全不同的策略。例如，当温度突然升高时，细胞会合成某些特殊蛋白质即热激蛋白（heat shock protein）。在各种不同生物中，许多热激蛋白非常相似。如大肠杆菌的 66kd 的热激蛋白 Dnak 就有一半序列与果蝇的热激蛋白 Hsp70（70kd）相同。大肠杆菌中某些热激蛋白还是噬菌体生长所必需的，如 GrpEL 和 GroES 参与 λ 粒子的组装过程，Dnak 参与 λDNA 的合成。热激蛋白的合成还与其他一些因素有关，如在大肠杆菌培养液中加入乙醇，也会诱导细菌合成热激蛋白。

热激的效应是使基因的表达方式突然改变。大肠杆菌受热激时，一种含有特殊 σ 因子的 RNA 聚合酶全酶识别热激蛋白基因的启动子。这种特殊 σ 因子的分子量为 32kd，是细胞内一种微量蛋白，其含量比普通 σ 因子（σ^{70}）低得多。含有 σ^{70}的全酶可转录大多数操纵元，但含有 σ^{32}因子的全酶只识别热激基因的启动子而不识别其他基因的启动子。热激蛋白基因的启动子不同于其它基因的标准启动子，其 -10 区和 -35 区的共有序列分别为 CCCCAT - T 和 T - C - C - CTTGAA。

虽然目前知道 σ^{32}能够特异性地识别热激蛋白基因的启动子，但对温度突然改变如何影响 σ^{32}因子调节热激蛋白基因表达的确切机制还不清楚。

小　结

原核生物的基因组为环状双链 DNA 分子，为负超螺旋结构。细菌染色体 DNA 只有一个复制起点和一个复制终点，为单复制子。细菌 DNA 复制涉及一种以上 DNA 聚合酶，大肠杆菌的 DNA 聚合酶Ⅲ合成前导链和后随链，DNA 聚合酶Ⅰ参与 DNA 修复过程。DNA 聚合酶Ⅲ的活性需要多种蛋白质的参与和需要复制体中（即复制启始区中）几段结构保守序列。DnaB 蛋白是一种参与复制叉形成的螺旋酶，它可能通过与 DnaG 引物酶互作，启始引物 RNA 合成。

DNA 复制启始时首先须使 DNA 双螺旋局部熔解，然后通过解螺旋形成单链。大肠杆菌 DNA 复制时，先后有几种蛋白质参与这一过程。DnaA 蛋白首先与 OriC 中一组 9bp 的重复序列和一组 13bp 的重复序列结合，使 13bp 的重复序列熔解。然后 DnaB 同 DnaC 一起将 DNA 进一步解螺旋。

大肠杆菌 DNA 复制时，引物体沿后随链的模板链移动，以相反方向合成引物。在 DNA 合成的延长期间，由引物体和 RNA 聚合酶Ⅲ全酶组成复制体。柯恩伯格模型认为，复制体有 2 个 DNA 聚合酶活性中心，它同时复制 2 条单链。其中一个活性中心连续合成前导链，另一活性中心间断地合成后随链而无须离开复制叉。

原核生物的染色体只有一个复制子，当两个复制叉在终止位点会合时，终止蛋白与终止位点进行特异互作，终止复制过程。

细菌细胞除了含有一个主染色体外，还有若干叫做质粒的小环状双链DNA。质粒也只有一个复制启始区，单拷贝质粒在每个细胞世代中只复制一次，多拷贝质粒在每个细胞世代中可以进行多次复制。质粒拷贝数由质粒的复制启始过程决定。

原核细胞只含有一种RNA聚合酶，转录所有基因。RNA聚合酶全酶可以分成两大部分，一种成分是由多亚基蛋白（$\alpha_2\beta\beta'$）组成的核心酶，它与RNA链的延长有关；另一部分为单亚基的σ因子，它与RNA聚合酶识别启动子和启始转录过程有关。

细菌启动子含有两段短的结构保守序列，二者分别位于转录起点上游－10位和－35位。RNA聚合酶首先与－35位的序列结合，然后再覆盖－10位的序列。最初的转录复合体为双链闭合复合体，然后－10区到转录起点之间的一段12bp的序列熔解，形成所谓开放复合体。转录启始时常常会发生流产转录启始，即RNA聚合酶合成一段2至9个碱基的RNA链后，便将这段RNA从复合体上释放下来而不离开启动子。在这一阶段的后期，σ因子从复合体上释放下来，然后核心酶沿DNA链移动，继续合成RNA。当RNA聚合酶到达终止位点时，聚合酶和RNA便从复合体上释放下来，然后DNA又完全恢复到双螺旋构型。

启动子的强度与－35和－10位的同感序列的同源性有关，同源程度愈高，启动子强度愈大，RNA聚合酶启始转录的效率愈高。但是启动子强度也受转录起点上游和下游序列的影响。RNA聚合酶通过使用不同σ因子识别具不同同感序列的启动子。

RNA聚合酶在两种不同终止位点上终止转录。一处为内向终止子，含有一个富含GC的发夹和一串碱基U。另一处为依赖于ρ因子的终止子，富含碱基C而少碱基G。原核生物基因转录的终止过程受两种机制调节，一种是某种蛋白质与转录复合体互作，或者促进终止，或者形成反终止子，继续转录下游顺反子。另一种是RNA构型在形成终止子和反终止之间互变。

细菌中编码蛋白质的基因的表达通常是转录和翻译过程偶联在一起，即在上游顺反子进行翻译的，同时，下游顺反子进行转录。在翻译启始期，mRNA和启始tRNA与核糖体结合，三种启始因子IF_1、IF_2、IF_3按步骤结合进来，形成启始复合体。AUG通常是翻译的起点密码。原核生物的大多数mRNA在起点密码的5′端都有一段与核糖体中16SrRNA互补的SD序列，二者通过碱基互补配对使核糖体与mRNA的翻译启始区结合。释译启始效率也受起点密

码下游的碱基的影响和 mRNA 次级结构的影响。翻译延长和终止过程在原核生物和真核生物中都基本相似。原核生物的 mRNA 大多为多顺反子 mRNA，即一条 mRNA 分子编码出几种不同蛋白质，但是各种蛋白质并不一定等量合成，有些 mRNA 的上游顺反子的翻译水平高于下游顺反子，有些则下游顺反子翻译水平高于上游顺反子。

原核生物基因的表达受细胞内外各种因素的调节，根据操纵元在通常情况下表达与否，可以将其分成负控制型操纵元和正控制操纵元，前者通常表达，而后者则通常不表达。所以负控制型又叫做抑制型，正控制型又叫做诱导型。

RNA 聚合酶在启动子上启始转录的功能既可受某些蛋白质的活化，也可受某些蛋白质的抑制。抑制蛋白或者阻止 RNA 聚合酶与启动子结合，或者抑制其启始转录的功能。抑制蛋白的结合位点叫操作子，通常位于转录起点周围或上游。抑制蛋白与操作子结合受某种小分子化合物的调节。小分子诱导物阻止抑制蛋白与操作子结合，辅抑制物则活化这一结合过程。

乳糖操纵元通过诱导物 β－半乳糖苷阻止抑制物与操作子结合进行表达，其 lacZ 基因通过转录和翻译产生一种水解 β－半乳糖苷的 β－半乳糖苷酶。色氨酸酶操纵元则通过辅抑制物色氨酸活化抑制蛋白，使抑制物与操作子结合，阻止与编码色氨酸合成有关的酶的基因表达。

RNA 聚合酶通常不能识别正控制型操纵元的启动子，只有存在某种特殊活化蛋白时，RNA 聚合酶才能转录这类操纵元。正活化蛋白也可能受某种小分子化合物的调节，如 CAP 活化蛋白只有与 cAMP 结合后才能结合到靶序列上。

基因表达过程可以在翻译水平上进行控制。某些蛋白质可以同 mRNA 上所谓翻译操作子结合，阻止核糖体启始翻译过程。还有一些操纵元的表达可以在翻译水平上进行自动调节，如各种核糖体蛋白操纵元都有一个顺反子编码与自身操纵元结合的蛋白质，通过这种蛋白质调节自身操纵元的表达。

细菌可以通过一种调节翻译终止过程的、叫做弱化反应的机制控制某些操纵元的表达。这类操纵元主要是一些与氨基酸合成有关的酶的操纵元如色氨酸操纵元等。在这类操纵元的 mRNA 的 5′端都有一段序列能形成几种可变的次级结构。究竟形成何种结构，由 mRNA5′端一段短的前导序列是否进行翻译所决定。在存在某种负载特殊氨基酸的酰基 tRNA 的条件下，如色氨酸操纵元的色氨酸氨酰基 tRNA，核糖体翻译前导序列，在前导序列下游形成终止子，下游顺反子不能转录。如果不存在这种氨酰基 tRNA，核糖体滞留在前导序列上，便形成另一种次级结构即反终止子而不形成终止子，从而使下游顺反子进行转录。在这种调节机制中，氨酰基 tRNA 的存在与否控制着氨基酸的生物合

成。

一种叫做反义 RNA 的小分子 RNA 也可以控制某些操纵元的表达。反义 RNA 能够与 mRNA 上的靶序列进行互补，形成双链 RNA 区，将翻译起点遮盖起来，直接或间接地阻止翻译。

原核生物的基因表达调控系统除了采用某些特殊机制对一些特定的基因表达进行控制以外，还可以从总体水平上对某些环境条件作出反应。当细菌在缺少营养的培养基上生长时，一方面中止 rRNA、tRNA、核糖体蛋白、糖类和脂肪的合成，另一方面则提高某些氨基酸合成酶基因的表达水平而表现出所谓严紧反应。严紧反应受 ppGpp 和 pppGpp 调节。当细菌在温度突然升高的条件下生长时，RNA 聚合酶的核心酶同 σ^{32} 结合，转录热激蛋白基因，合成一组热激蛋白。

复　习　题

选择 1～3 题中正确答案：

1. 复制体的组成包括以下哪些因素？

A. 一种在启始位点上与 DnaG 引物酶互作的寡聚体酶

B. 一种阻止 DNA 降解的单链结合蛋白

C. DnaB 螺旋酶以及一些辅助蛋白如 DnaC、DnaT 等

D. DnaB、SSB 蛋白、DnaC、DnaT、DnaG

E. DnaB、DnaG、DNA 聚合酶Ⅲ全酶。

2. 原核生物中，操纵元的转录可以在下列哪一水平上进行调节？

A. RNA 聚合酶全酶补给到 DNA 上（σ 因子与 DNA 特异结合）

B. 抑制蛋白或活化蛋白结合到启动子附近

C. 转录终止和反终止

D. 以上所有水平

3. 以下哪些可以说明某个原核生物操纵元的基本特征，操纵元

A. 通常由多基因组成，各个基因都位于一个启动子和一个终止子之间

B. 可以在转录水平上自动控制

C. 操纵元中各个基因的上游都有相同的启动子·

D. 各个基因的表达是并列控制的

4. 区别大肠杆菌的引物酶、单链 DNA 结合蛋白、螺旋酶、DNA 连接酶、DNA 聚合酶Ⅰ和 DNA 聚合酶Ⅲ在 DNA 复制过程中的作用。

5. 大肠杆菌 DNA 在复制过程中复制叉每秒钟向前移动 500bp，那么复制

叉之前的 DNA 每秒需转动多少次？

6. 乳糖操纵元的编码产物之一是 β－半乳糖苷酶，该酶催化乳糖分解成半乳糖和葡萄糖，乳糖的半乳糖苷部分具有诱导功能和通过与 lac 抑制子互作，解除其对乳糖操纵元的抑制作用。请说明为什么可用 x－gal 和 IPTG 来检测半乳糖苷酶的存在与否？

7. 对照大肠杆菌启动子中－10 框的同感序列，（1）在 CATAGT 这段序列的第一个碱基发生 C→T 突变，这种突变是加强突变还是削弱突变？（2）如果最后一位碱基发生 T→A 突变，这种突变是加强突变还是削弱突变？

8. 下表是利用 1 个和 2 个拷贝的 lac 操纵元测定大肠杆菌细胞中 β－半乳糖苷酶和乳糖渗透酶的结果。其中一个拷贝位于细菌染色体上，用黑体表示；另一拷贝位于 F 质粒（F 因子）上，用斜体表示。根据表中基因型，用“－”或“＋”填空，以说明这两种酶的合成情况。

菌株	基　因　型	β－半乳糖苷酶		渗透酶	
		－IPTG	＋IPTG	－IPTG	＋IPTG
1	$\mathbf{I^+P^+O^+Z^+Y^+}$				
2	$\mathbf{I^+P^+O^cZ^+Y^+}$				
3	$\mathbf{I^+P^+O^cZ^+Y^+}/\mathit{I^+P^+O^+Z^+Y^+}$				
4	$\mathbf{I^+P^+O^cZ^-Y^+}$				
5	$\mathbf{I^+P^+O^cZ^-Y^+}/\mathit{I^+P^+O^+Z^+Y^+}$				
6	$\mathbf{I^+P^+O^cZ^+Y^-}$				
7	$\mathbf{I^+P^+O^cZ^+Y^-}/\mathit{I^+P^+O^cZ^+Y^+}$				
8	$\mathbf{I^-P^+O^+Z^+Y^+}$				
9	$\mathbf{I^-P^+O^+Z^-Y^+}/\mathit{I^+P^+O^+Z^+Y^+}$				
10	$\mathbf{I^-P^+O^+Z^+Y^+}/\mathit{I^+P^+O^+Z^-Y^+}$				
11	$\mathbf{I^-P^+O^+Z^+Y^+}/\mathit{I^+P^+O^+Z^+Y^-}$				
12	$\mathbf{I^+P^-O^+Z^+Y^+}/\mathit{I^+P^+O^+Z^+Y^-}$				

9. 当培养基中色氨酸浓度从高含量降到极低含量时，以下大肠杆菌细胞中色氨酸操纵元的转录水平可能会提高或者会降低？为什么？

A. 野生型

B. 前导序列中 2 个色氨酸密码从 UGGUGG 突变成 UGAUGG（UGA 为终止密码）的突变体。

C. 前导序列中色氨酸密码从 UGGUGG 突变成 UCGUGC 的突变体。

10. 以抗红霉素甲基化酶 mRNA 的翻译为例，说明原核生物 mRNA 翻译的自动调节机制。

11. 比较操纵元正控制表达和负控制表达的基本区别。

第四章

真核生物细胞核基因组及其表达

真核细胞区别于原核细胞主要在于有细胞核，核内遗传物质由核膜将其与细胞的其他成分隔开。真核细胞的遗传物质主要存在于细胞核内，细胞内的其他细胞器如线粒体和叶绿体等也含有少量遗传物质。真核细胞与原核细胞的另一重要差别是细胞核 DNA 通常与组蛋白结合，形成细胞分裂间期的染色质（chromatin）或细胞分裂周期中的染色体（chromosome）。细胞核 DNA 以线状染色体的形式存在，而不是环状染色体。每个细胞中染色体的数目和大小因物种不同变化很大。由于细胞核具有核膜，所以其遗传信息的转录和翻译就由核膜隔开，分别在细胞核和细胞质中进行。本章将主要讨论细胞核 DNA 的复制、基因的结构、组织与表达特点，有关细胞质中线粒体和叶绿体基因组及其与核基因组之间的互作将分别在第五章中讨论。

第一节　DNA 与染色体

一、组蛋白与染色质

在真核细胞中，每一条染色体都由两条互补的、反向平行的 DNA 单链组成，其线状 DNA 通常都与若干种蛋白质结合。在体细胞中，染色体大多成对存在，只是多细胞动物和某些植物的性染色体没有同源成员，成单存在。

在有性生殖过程中，精、卵细胞都只含有一条来自其亲代的染色体，所以受精以后，合子（zygote）就含有两条同源染色体，它们分别来自父、母双方。真核生物都由一个受精卵经过多次一分为二的细胞分裂发育而成，如果不出现自发突变，同一个体的各个细胞的遗传组成都相同。

细胞核中的染色体由 DNA 和蛋白质组成。与细胞核 DNA 结合的蛋白质叫组蛋白（histone），组蛋白有 5 种主要类型（表 4－1），普遍存在于各种动、植物细胞核之中。这 5 种组蛋白的分子量都较小，其中带正电荷的氨基酸（精氨酸和赖氨酸）多于带负电荷的氨基酸（谷氨酰胺和天冬酰胺），所以，各种

组蛋白非常适合与带负电荷的核酸结合。在某些类型的细胞中存在另一种组蛋白 H_5，如鸟类和两栖动物的红血细胞中，H_5 代替 H_1 将 DNA 紧密地包裹起来。鱼精细胞比较特殊，含有鱼精蛋白（protamine）而不是组蛋白。鱼精蛋白为一组高度带正电荷的蛋白质，这种蛋白质可能适合将精子头部的 DNA 包裹成很小的体积。

在不同生物中，各种组蛋白都是高度保守的。例如在奶牛和百合花中，H_4 组蛋白只有 2 个氨基酸的差别。这种现象表明，组蛋白结构的高度保守性对其特殊功能是极为重要的。

真核细胞中还有其他一些蛋白质与 DNA 结合，由于其含量很低，且难以分离，所以其功能目前还不清楚，可能其中有些蛋白质对控制转录具有重要作用。

表 4－1 5 种主要的组蛋白

符 号	氨基酸数	正电荷氨基酸%	
		精氨酸	赖氨酸
H_1	215	1.4	28.4
H_2A	132	9.1	9.1
H_2B	124	6.5	16.1
H_3	135	13.3	8.9
H_4	102	13.7	9.8

组蛋白的主要功能是将 DNA 包裹成一种特殊的、叫做核小体（nucleosome）的螺旋结构。核小体是染色质的基本单位，其核心是由 4 种组蛋白各以 2 个分子组成的八聚体（H_2A、H_2B、H_3、H_4），DNA 分子则用大约 140 对碱基绕八聚体核心 2 圈，形成核小体（图 4－1）。组蛋白核心的中央为由 H_3 和 H_4 组成的四聚体，两侧分别为一个 H_2A-H_2B 二聚体。两个相邻核小体由大约 60bp 的 DNA 相连，这段 DNA 叫做连接丝，一般由组蛋白 H_1 包裹。H_1 的氨基末端部分结合到八聚体的外围，其羧基末端部分则与连接丝中的 DNA 互作。连接丝 DNA 的长度变动于 60～80bp 之间。核小体和连接丝组成染色质的基本单位核小体重复（nucleosome repeat），一个核小体重复含有约 200 对碱基。染色质还可进一步螺旋化，形成更为致密的结构如直径为 30nm 的螺旋体（solenoid）或光学显微镜下见到的染色体。核小体重复也可以通过各种步骤而解体，如除掉 H_1 等。一般认为，这一过程与基因转录有关。H_1 对染色质收缩

成超螺旋体起着关键的作用，芽殖酵母（*Saccharomyces cerevisiae*）不含 H_1，所以在有丝分裂期间很难见到染色体。

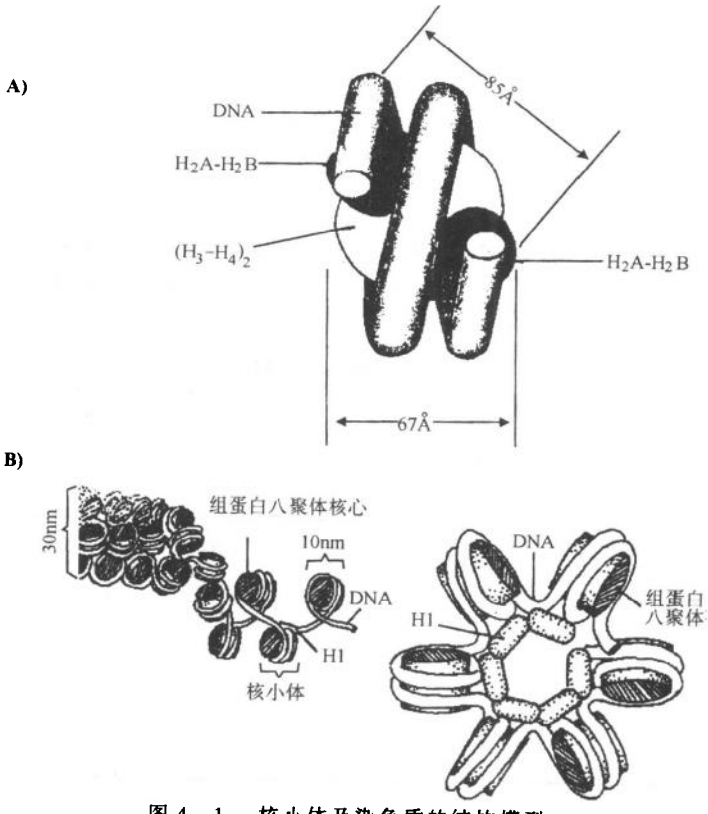


图 4－1　核小体及染色质的结构模型

A. 示 DNA 绕组蛋白八聚体核心 2 圈　B. 超螺旋的结构模型，左示部分解螺旋的 30nm 的超螺旋体，右示，从末端观察超螺旋体

依据间期细胞核中染色质的浓缩状态，可以将染色质分成两大类。在细胞分裂间期，大多数染色体物质都解除浓缩，染色较浅，这类染色质叫做常染色质（euchromatin）。但是，染色体的某些区域此时并不松散，仍保持其高度螺旋状态，所以在整个间期都具有很强的染色特性，这类染色质叫做异染色质（heterochromatin）。异染色质和常染色质除了在染色特性方面存在差别外，在其他方面也存在很大差别。首先，异染色质在 RNA 合成方面活性较低；第二，异染色质是细胞分裂过程中最晚复制的 DNA；第三，在减数分裂期间，异染色质抑制非姊妹染色单体之间的交换。

异染色质也有两类即组成型异染色质（constitutive heterochromatin）和兼

性异染色质（facultative heterochromatin）。组成型异染色质始终可以在两条同源染色体的同一区域见到，大多位于着丝粒（centromere）区，通常由高度重复的DNA序列组成。而兼性异染色质只是在生物体的某个特殊发育阶段，染色质通过浓缩而形成，通常只出现在两条同源染色体之一的某个区域，该区域的DNA通过浓缩成异染色质而不具有转录活性。

二、着丝粒与卫星DNA

真核细胞中，各条染色体的两条臂之间都有一个特殊的、叫做着丝粒的结构，而着丝粒所在的染色体区都由若干种高度重复的DNA组成。这些重复DNA都是一些简单序列，可以重复数千次，甚至更高。由于这类重复序列的组成同细胞中其它大多数DNA序列的组成不同，所以通过氯化铯（cesium chloride）梯度离心DNA片段，很容易将其同其它DNA分离开来，成为DNA主带的一种附属带。这种DNA叫做卫星DNA（satellite DNA）。着丝粒区含有三种卫星DNA，三者都由高度相似的序列组成，第一种含有ACAAACT基本重复，第二种含有ATAAACT重复，第三种含有ACAAATT重复。

卫星DNA通常不转录，因为其有限的DNA序列缺少启动子；但是卫星DNA含有能被DNA聚合酶识别的核苷酸序列。由于着丝粒DNA能够迅速复制，可能着丝粒附近的卫星DNA与着丝粒特异性蛋白的结合有关，因此它只是真核生物基因组中一种结构区域，而不是遗传区域。

在酵母细胞中，每一条染色体都只有一个着丝点，都只通过着丝点与一条纺锤丝相连，也都只有一段特殊的着丝点序列（centromeric sequence）。根据对芽殖酵母的研究，所有染色体的着丝点序列都位于长130bp的、富含碱基AT的区域之中，都具有2段短的、高度保守的序列(图4-2)。

	因子Ⅰ	因子Ⅱ	因子Ⅲ
CEN3	ATAAGTC	ACATGAT←88bp（93%AT）→	TGATTTCCGAA
CEN11	ATAAAGTC	ACATGAT←89bp（94%AT）→	TGATTTCCGAA
CEN4	AAAGGTC	ACATGCT←82bp（93%AT）→	TGATTACCGAA

图4-2 啤酒酵母三个着丝点（cen）序列中具有同源性的区域

因子Ⅰ和因子Ⅲ高度同源，二者之间由一段较长的、富含碱基AT的序列（因子Ⅱ）相连

由于这种着丝点序列对核酸酶的抗性较强，可能它与某些非组蛋白结合，调节染色体与纺锤体的微管结合。对于高等真核生物来说，由于每一条染色体都与许多条纺锤丝相连，因此着丝粒区中卫星DNA序列的高度重复性就是产生这

种现象的基础。由于酵母的每条染色体都只与一条纺缍丝相连，所以其着丝粒区不存在重复 DNA。

三、端　　粒

真核细胞的 DNA 为线状分子，其染色体末端结构非常特殊，这种染色体的末端结构叫做端粒（telomere）。根据对原生动物尖毛虫（*Oxytracha*）端粒 DNA 序列的研究，端粒 DNA 折叠成一种四聚体结构，形成一种核蛋白复合体，外由端粒蛋白包裹。在纤毛虫（*ciliata*）中，现已发现有两种端粒蛋白，已知其中分子量为 55kd 的蛋白质的氨基酸序列类似组蛋白 H_1。端粒蛋白就像一顶保护端粒的帽子，防止染色体末端被降解、发生融合和重组。

端粒 DNA 由许多短的、首尾连接的重复序列组成，其中一条单链的 3′末端区富含碱基 G，叫做 G 链，另一条与其互补的链叫做 C 链。在各种不同的真核细胞中，端粒 DNA 序列及其结构都很相似。例如在人体以及其他许多脊椎动物染色体的末端，G 链都含有 2～30kb 的端粒重复 TTAGGG 序列，这段序列按 5′→3′方向延伸，并超出其互补 C 链 12～16 个核苷酸（图 4－3）。这些富含碱基 G 的单链重复能够通过形成 G—G 碱基对而折叠起来，但目前在活体内并未测定到这种结构。在端粒序列之中或附近，还存在另一些与端粒形成有关的序列，它们广泛分布于从酵母到人体各种不同生物的染色体端粒之中，但其长度和组成具有很强的物种特异性。

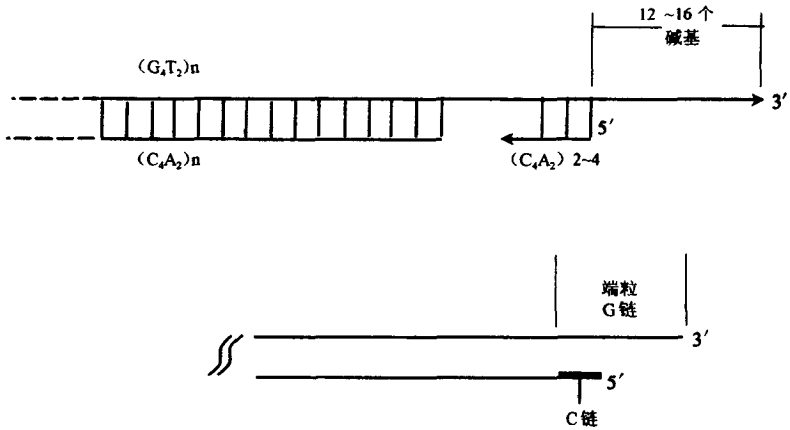


图 4－3　真核细胞的染色体端粒

下图为端粒 DNA 的 C 链和 G 链；上图为四膜虫端粒的末端 DNA 结构

在酵母中，每一条染色体的单粒也都由端粒重复组成，总长度约100bp，其中一条单链的3′末端的端粒重复序列为TG_{1-3}。同高等真核生物一样，酵母的端粒DNA也不是由常规的DNA复制机器合成，也都是由酶将这种序列加到DNA链的3′端，然后再通过一至数种类似引物酶的酶合成与其互补的$C_{1-3}A$畜列。在酵母染色体的内部也存在一些$C_{1-3}A$序列，所以酵母染色体在$C_{1-3}A$重复序列之内或在其附近断裂时，通过重组修复，可以获得新的端粒序列。

四、分散的高度重复DNA

几乎所有高等真核生物的基因组都含有不集中排列在一起的高度重复DNA序列，它们比较均匀地分散在整个基因组中，由一些单拷贝DNA或中度重复的DNA隔开。

在人体基因组中，大多数分散排列的重复DNA都属于Alu家族，其每一成员长约300bp，序列比较接近，但并不完全相同。由于这类DNA序列的中部都有一个限制性内切酶AluI的切割位点，所以命名为Alu序列（Alu sequense)。人体基因组中的Alu序列约有10万个拷贝，约占总体DNA的3%～6%，相当于每5kb或稍长一些的DNA片段就可能含有一段Alu序列。显然，很多基因的内含子或与基因相邻的序列都含有Alu序列。

一般认为，人体基因组中Alu序列的广泛分布与调节RNA合成有关。Alu序列的3’区具有富含碱基A的重复尾巴，其两侧通常接有8～19bp的顺式重复。在离体条件下Alu序列很容易由RNA聚合酶Ⅲ转录，但在活体内转录水平并不与其拷贝数成比例。

在其他哺乳动物的基因组中也都存在与人体Alu序列相似的序列，而且这些Alu序列及其类似序列都与另一种叫做7SL的小RNA分子存在同源性。7SL RNA长300nt，与6种蛋白质结合，形成一种细胞质RNA－蛋白质复合体称为SRP。SRP对新合成的蛋白质穿越内质网的移位过程具有重要作用（见本章第四节)。在人体基因组中，7SL RNA由RNA聚合酶Ⅲ从3～4个相同拷贝的基因转录而成，其5′端和3′端序列都与Alu序列存在同源性，中间155个碱基的序列不同源，可能Alu序列起源于7SL RNA或其DNA。

在哺乳动物和其他高等真核生物基因组中除了存在Alu家族的成员以外，还存在其他一些广泛分布的高度重复序列，其中有些很短，约500bp，两侧接有顺式重复，与tRNA存在同源性，它们可能起源于tRNA的中间产物。还有

一些高度重复的序列很长（5kb 以上）具有富含碱基 A 的 3′序列，其中有些还含有可读框。因此，对真核生物基因组中的转座子、假基因和重复序列很难加以区分。

五、细胞分裂中期的骨架

关于高等真核细胞间期的染色质形成细胞分裂中期染色体的机制目前不清楚，但是现在知道，有丝分裂中期的染色体是由许多染色质环附着到一种蛋白质骨架上形成的，这种骨架叫做细胞分裂的中期骨架（metaphase scaffold）。例如，用肝素（heparin）或硫酸葡聚糖（dextran sulfate）处理细胞分裂的中期染色体时，可以除掉组蛋白和非组蛋白，这种除掉了蛋白质的染色体仍保留了有丝分裂中期的形态，但都由一种未折叠的 DNA 晕圈所环绕。DNA 晕圈由许多 DNA 环组成，每一个环的长度为 10～30μm（35～100kb），都在其基部固定到中期骨架上。除掉了组蛋白的中期骨架主要由两种高分子量的蛋白质（170 和 135kb）组成，这两种蛋白质可能形成一种网络，作为将所有 DNA 环连接在一起的中轴。有关中期骨架的组装过程目前还不清楚，如果 170kD 的蛋白质是拓扑异构酶Ⅱ（本章第二节），那么 DNA 超螺旋的变化可能是中期骨架形成的关键。

第二节　细胞核 DNA 复制

一、复制子及复制启始区

现在知道，在各种真核生物中，每一条染色体都是从多个复制起点开始，两个复制叉分别朝两个方向延伸的。因此每一条新合成的 DNA 片段就叫做一个复制子（replicon）。其长度等于两个相邻复制起点之间的距离（图 4－4）。这个复制起点就叫做复制启始区（replication origin）。在高等植物中，双子叶植物的复制子长 20～30μm（70～100kb），单子叶植物的复制子一般长 5～21μm。在哺乳动物中，两个相邻复制启始区约相距 50～100kb，非常接近细胞分裂中期染色体的 DNA 环的长度，因此各个复制单位或复制子可能相当于中期骨架上两个相邻附着点之间的 DNA 超螺旋部分。DNA 复制时，核小体并不从中间分裂，使 2 个子双螺旋各得半个核小体，而是其中一条子双螺旋接收一个旧的核小体，另一条子双螺旋则获得一个新的核小体。

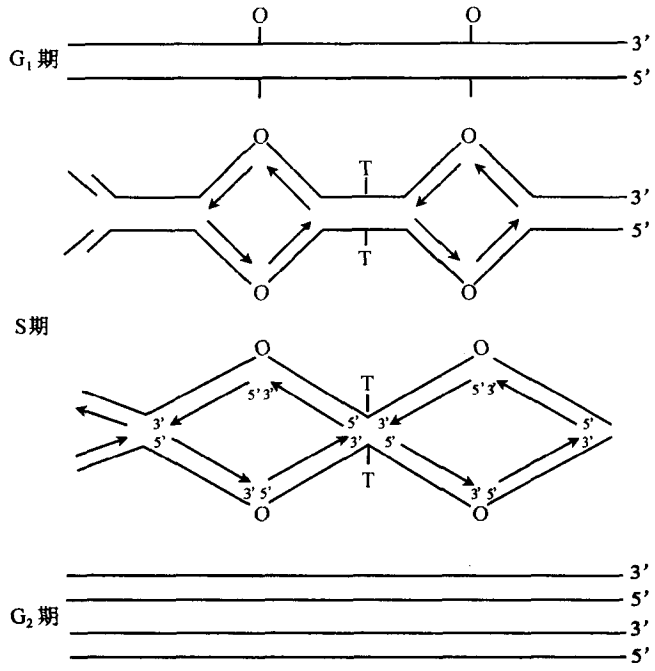


图 4－4　细胞周期中的复制子及 DNA 合成过程

O 为复制起点，T 复制终点，箭头示新链合成的方向（5′→3′），G_2 期中的 2 条双链 DNA 表示细胞分裂时的两条姊妹染色单体

根据对 SV_{40} 和动物多瘤病毒的研究，复制启始区含有一段短的、富含碱基 AT 的顺接重复序列 TTAACGTTAA，由于这段序列为一种双向对称结构，即分别从两条单链的 5′至 3′方向阅读时，其序列相同。所以，这种 DNA 在复制过程中有可能形成发夹结构，作为启始复制过程的一种信号。目前还不清楚在不同真核生物中细胞核 DNA 复制启始区的结构是否都相似。

在 S 期，并非所有复制启始区都同时进行活动，而是一些相邻的复制子在某个特定时间启始和终止复制过程。在哺乳动物细胞中，每一复制子簇约含 25～100 个复制子。根据对哺乳动物染色体分带的研究，当用 Giemsa 染料染色之后，可见到其染色体上约有 2 000 条明暗相间的带纹。这些带纹除了反映 DNA 分子中不同区域的碱基组成不同以外，还可反映出 DNA 浓缩程度的细微差异。间期染色质中，每一个协调定时复制的复制子簇都相当于中期染色体上某条特殊带纹。例如，在 S 期的前半期用 5－溴脱氧尿苷（BudR）处理细胞时，可以标记染色体上着色浅的带纹，在 S 期的后半期处理时，则可标记着色深的带纹。在 S 期的前期和后期之间，DNA 合成速率明显下降，形成一个低谷。现在知道，一些异染色质区包括着丝粒附近的区域以及雌性哺乳动物的不

具活性的X染色体都相当于着色深的染色体区，这些区域在S期的后半期复制。此外，在所有细胞中，那些组成型表达的基因（有时叫做看家基因）复制较早，而一些组织特异性表达的基因在不表达的细胞中通常复制较晚。因此，基因组中活动的基因可能与早期复制和着色浅的染色体带之间存在一定相关性。

二、DNA复制的蛋白质或酶类

细胞核DNA复制所需酶类或蛋白质的种类及其确切的作用机理目前都还不清楚，下面就目前已知的几种复制酶作扼要说明。

1.单链DNA结合蛋白（SSB）　细胞核DNA的复制启始区富含AT碱基，根据对小牛胸腺细胞中单链DNA结合蛋白的分析，这种蛋白质趋于同复制启始区中富含AT碱基的单链DNA结合，能激活α－DNA聚合酶，可能SSB蛋白在DNA复制过程中起着固定复制叉中单链区的作用。

从真菌（*Ustilago*）中分离到的SSB蛋白分子量为20kb，在50℃条件下能降低多聚d（AT）/d（TA）的Tm值（DNA半变性时的温度值）。植物（*Lilium*）中的SSB蛋白目前只在胞母细胞中有发现，当发生遗传重组时，其浓度达到高峰，可能SSB蛋白与复制和重组过程都有关系。在大鼠肝细胞瘤（Novikoff hepatoma）细胞中，SSB能激活β－DNA聚合酶，并能提高DNA合成速率。β－DNA聚合酶与DNA修复过程有关，可能SSB蛋白通过改变模板链的构型，以促进DNA的修复合成和DNA复制。

2.DNA引物酶　同原核细胞DNA复制一样，细胞核DNA复制也需要由引物酶首先合成引物RNA。在果蝇和哺乳动物胚胎中，α－DNA聚合酶的亚单位通常与引物酶结合，形成DNA聚合酶—引物酶复合体。这种复合体至少由三个亚基组成，即α、β和γ亚基，其分子量分别为182kd、60kd和50kd，其中α亚基即α－DNA聚合酶与DNA引物酶的活性有关。近年来，从小麦胚胎的α－DNA聚合酶中也分离到一种DNA引物酶，其分子量为70kd。

3.DNA聚合酶　在高等真核细胞中存在多种DNA聚合酶即α、β、γ、δ和ε，高等植物中α－DNA聚合酶（α－DNA polymerase）可能为复制酶。一般说来，α－DNA聚合酶是一种高分子量的蛋白质，在菠菜的叶片中，其分子量为160kd，由2个分子量均为80kd的亚基组成，但其活性部分的分子量为12kd。在豌豆中，α－DNA聚合酶中具有催化活性的亚基的分子量为180kd甚至更高，该亚基可以降解成一系列分子量不同的亚基（140、110、75和45kd）。豌豆中的这种酶至少还具有另外2个非催化活性的亚基。α－DNA聚

合酶对低浓度的 N－乙顺丁烯二酰亚胺（N－ethylmaleimine，NEM）敏感，特别是对动物 α－DNA 聚合酶的一种特殊抑制剂蚜栖菌素（aphidicolin）敏感。植物中大多数 α－DNA 聚合酶都不具有核酸酶活性。根据对哺乳动物各种 DNA 聚合酶的研究，一般认为 α－聚合酶与后随链合成有关，δ－聚合酶（也有可能是 ε）与前导链合成有关。这两种 DNA 聚合酶在持续合成能力即在执行复制任务的持续时间方面有很大差别，δ－聚合酶与一种叫做增殖细胞核抗原（proliferating cell nuclear antigen，PCNA）结合，其持续合成能力比 α 聚合酶大得多。PCNA 是 DNA 复制非常活跃的增殖细胞中一种含量丰富的蛋白质，可使 δ－聚合酶的持续合成能力提高 40 倍，即它使 δ－聚合酶延长 DNA 链的长度提高 40 倍。

在小麦中 γ－DNA 聚合酶可能也是一种复制酶，由 3 个亚基组成，并具有识别引物的能力。哺乳动物 γ－DNA 聚合酶存在于线粒体中，它与线粒体 DNA 复制有关，而 β－聚合酶可能是一种 DNA 修复酶。

4. *核酸酶 H*　RNA 酶 H 是一种特别能降解 DNA－RNA 杂种分子中 RNA 的酶，其分子量为 70～150kd，在存在二价阳离子的条件下具有活性，为一种内切核酸酶。虽然在 DNA 合成过程中该酶活性增加，但目前尚无直接证据说明该酶与复制有关。

5. *DNA 连接酶*　DNA 连接酶催化两条相邻 DNA 片段的 3’－OH 和 5’磷酸基之间形成磷酸二酯键。在原核生物中该酶依赖于 NAD，而在真核细胞中则依赖于 ATP，并同时需要 ATP 和 Mg^{++}。在植物中该酶的活性机制还不清楚。

6. *螺旋酶*　螺旋酶是一种依赖于 DNA 的 ATP 酶，它利用 ATP 打断碱基对之间的氢键，因此其功能是使双螺旋解链。在百合属植物中，该酶的分子量为 130kd，但其确切功能目前不清楚。

7. *回旋酶和拓扑异构酶*　在原核生物中，回旋酶将 DNA 超螺旋解螺旋成一种适合于形成复制叉的双螺旋结构，并使其他几种酶结合到复制叉上。回旋酶是一种特殊的拓扑异构酶，拓扑异构酶的主要作用是使超螺旋 DNA 分子解螺旋。在真核生物中，目前尚未发现回旋酶，但发现另外两种拓扑异构酶。第一类拓扑异构酶（Top Ⅰ）通过切割双链中的一条单链，并使其旋转，然后再将缺口缝合起来而使超螺旋松弛。第二类酶（Top Ⅱ）通过使两条单链同时形成缺口，然后再将缺口缝合，使松弛的 DNA 再形成依赖于 ATP 的正超螺旋。一般认为，Top Ⅰ参与复制叉的形成过程。小麦胚胎中的 Top Ⅰ具有典型的真核生物特征，它不需要 Mg^{++} 和 ATP，能够使超螺旋松弛，其分子量为 110kD。豌豆中的这种酶也不需要 Mg^{++} 和 ATP，但需要 NaCl；此外，磷酸

根离子也有促进其活性的作用。

8. 内切脱氧核糖核酸酶　虽然一般认为拓扑异构酶与超螺旋的松弛过程有关，但某些研究表明，单链特异性的内切核酸酶也参与 DNA 复制过程。这类酶在双链 DNA 中富含碱基 AT 的区域切割，因此它可以在一定间隔距离内切割一条 DNA 单链，并使其自由转动。当复制叉通过这段序列以后，内切脱氧核糖核酸酶再将缺口修补好。在许多植物中都已发现内切脱氧核糖核酸酶，如豌豆的这种酶与染色质结合，只有一条多肽链，分子量为 18kD，与 DNA 复制有关。如果拓扑异构酶与超螺旋的松弛过程有关，那么内切核酸酶则与复制后的修复过程有关。

9. 端粒酶（telomerase）　端粒酶不直接参与 DNA 复制过程，但它与染色体端粒 DNA 合成有关。端粒酶为一种核蛋白复合体，由 RNA 和蛋白质两部分组成。就其功能而言，它是一种依赖于端粒特异性 RNA 的 DNA 聚合酶，是一种特殊的 RNA 逆转录酶，其 RNA 模板是该酶的组成成分。端粒酶中的 RNA 和蛋白质成分对端粒 DNA 合成都具有功能，其 RNA 具有很强的物种特异性。因此，一般认为端粒酶补偿细胞内普通复制机器不能复制线状 DNA 末端的缺陷。

三、细胞核 DNA 复制模型

虽然目前有关真核生物细胞核 DNA 复制的生物化学过程以及各种复制的作用机理都还不清楚，但是根据脊椎动物，无脊椎动物以及真核生物病毒的 DNA 复制系统的研究结果，一般认为细胞核 DNA 复制过程同大肠杆菌 DNA 复制既相似又有差别。

根据从哺乳动物细胞中分离到的各种 DNA 聚合酶和蛋白质离体条件下复制猴肾病毒 SV40DNA 的研究知道，复制首先由病毒基因编码的、叫做大 T 抗原（large T antigen）的位点特异性 DNA 结合蛋白结合到 SV40 惟一的一个复制启始区上开始（图 4－5）。大 T 抗原具多种功能（第六章），在这一过程中，它通过其螺旋酶活性将 DNA 双螺旋局部解螺旋，然后宿主细胞的叫做复制因子 A（replication factorA，RFA）的单链 DNA 结合蛋白固定分开的单链。

同原核生物 DNA 复制相似，先由引物酶合成一段引物 RNA，然后两个复制叉再进行双向复制。前导链的合成是连续的，而后随链的合成则是不连续的。与原核细胞 DNA 复制不同的是，真核细胞中有两种不同 DNA 聚合酶即聚合酶 α 和聚合酶 δ 作用于复制叉。聚合酶 δ 主要复制前导链，而与引物酶紧密结合的聚合酶 α 则合成后随链。

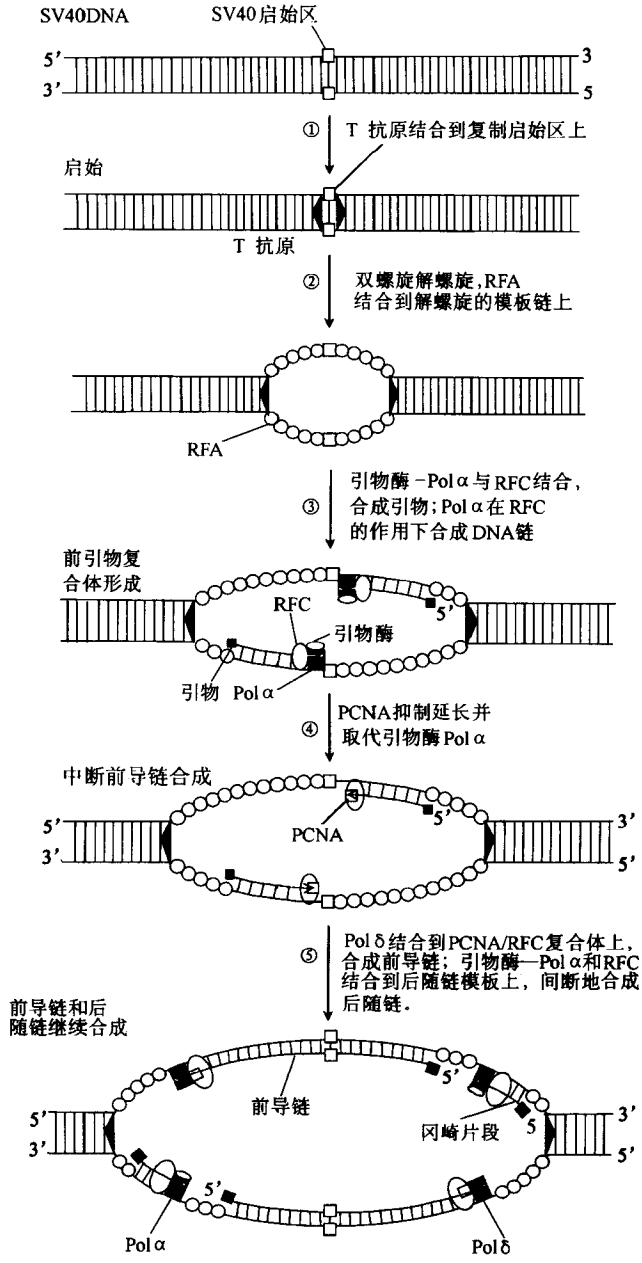


图 4－5　离体条件下真核细胞有关的酶复制 SV40 DNA 的模型

polα 和 polδ 分别为 DNA 聚 α 和 δ，RFA 和 C 分别为复制因子 A 和 C，PCNA 为增殖细胞核抗原

前导链最初合成时需要α-和δ-DNA聚合酶，当引物酶-聚合酶α复合体结合到解螺旋的模板链上后，由引物酶合成引物RNA，接着在复制因子C（RFC）的作用下，由α-DNA聚合酶合成一段短的DNA。然后增殖细胞核抗原PCNA结合到引物/模板的末端，替换α-DNA聚合酶，从而中断前导链合成。最后δ-DNA聚合酶再结合到增长链3'端的PCNA上，继续合成前导链。随着DNA双螺旋进一步解螺旋，引物酶/聚合酶α复合体又与前导链引物下游的解螺旋的模板链结合。接着，在两种DNA聚合酶和RFC的共同作用下，开始合成后随链，同时在复制启始区的另一侧继续合成前导链。随着复制叉不断向前移动，复制体又合成新的引物RNA和冈崎片段。当一个复制子的复制过程完成以后，两个相邻复制子上新形成的DNA链再连接起来。

四、染色体端粒DNA的合成

如本节前述，端粒是染色体末端的一种DNA-蛋白质复合体，端粒DNA在结构和功能方面与其他DNA序列显著不同，它由端粒酶通过一种特殊的方式合成（图4-6），所以在这个概念上，端粒合成又不同于通常的DNA复制过程。

例如在四膜虫（*Tetrahymena*）和游仆虫（*Euplotes*）中，端粒酶RNA序列为5'-CAACCCCCAA-3'，端粒酶以其自身的RNA为模板，以dGTP和dTTP为原料按5'→3'方向延长染色体末端。延长的G链中未配对的末端碱基再用于另一轮由端粒酶催化的延长过程，其他未配对部分则由引物酶和DNA聚合酶合成与其互补的C链。端粒酶按这种方式不断地将端粒重复DNA加到染色体末端，最后使染色体末端具有数百乃至数千对碱基的端粒重复，并使G链超出C链12-16个核苷酸。

五、有选择的DNA复制与基因扩增

真核生物在某一生长发育阶段，某些器官或组织可能对某些基因产物需求较多，才能满足其一定的代谢平衡关系，因此，生物体就有可能在其基因组中有选择地复制某个基因座位或某一段DNA序列，这种现象叫做基因扩增(gene amplification)。基因扩增主要见于昆虫、两栖动物以及植物的rRNA基因。某些昆虫如果蝇的卵壳蛋白（chorion）基因、哺乳动物的二氢叶酸还原酶基因等在某个发育阶段或某种特定环境条件下也都进行扩增。

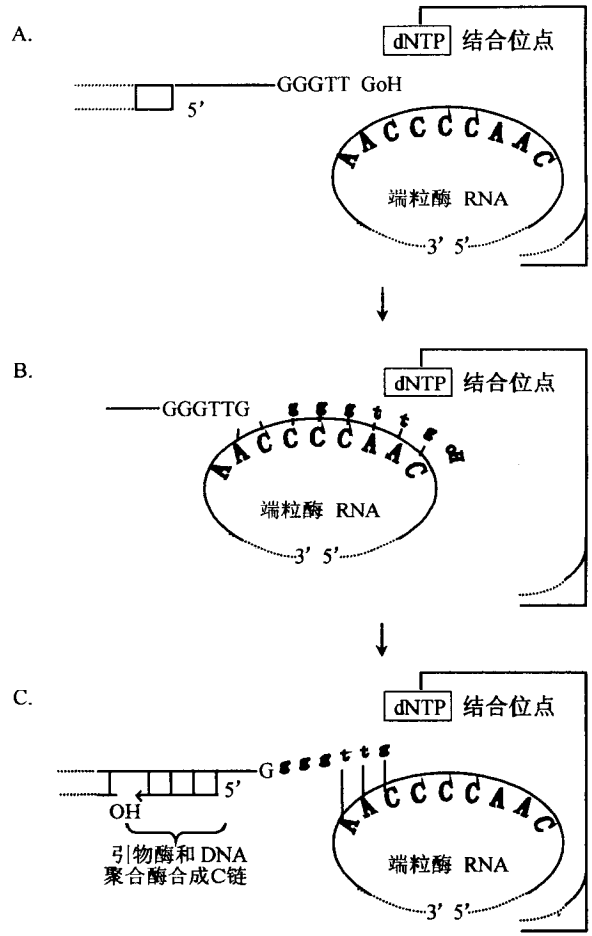


图 4－6　四膜虫中由端粒酶催化端粒 DNA 合成模型

A. 染色体末端 G 链向 3′ 方向突出的少数几个碱基与端粒酶 RNA 进行碱基互补配对 B. 以端粒酶 RNA 为模板，以 dGTP 和 dTTP 为原料延长染色体末端　C. 延长的 G 链中未配对的末端碱基再用于另一轮由端粒酶催化的延长过程，其他未配对部分则由引物酶和 DNA 聚合酶合成与之互补的 C 链

（一）rRNA 基因扩增

两栖动物的卵母细胞大约在减数分裂的第一次分裂的前期合成核糖体 rRNA，合成的高峰期是在同源染色体配对和复制完成以后的一段时间。在滑爪蟾（*Xenopuslaevis*）中，每一个二倍体细胞都含有 2 个 rRNA 基因的同源区，每一区域约含有 450 个拷贝的转录单位即 rRNA 操纵元。两个相邻功能单位之间都有一段非转录间隔区（图 4－7）。各个非转录区的长度各不相同，但都由许多短

的重复 DNA 组成；这些重复序列能与 RNA 聚合酶 I 结合，因此非转录区的功能可能是为前体 rRNA 的转录聚集转录酶。每一转录单位都转录出 18s、5.8s 和 28s 的前体 rRNA（图 4－7），前体 rRNA 然后通过加工才形成这三种 rRNA。核糖体中的 5SrRNA 则从基因组的另一位置转录而来，每个基因组大约有数千个拷贝的 5SrRNA 基因。

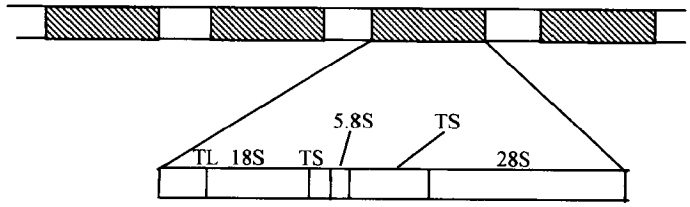


图 4－7　滑爪蟾中 rRNA 基因的组织

TL 为转录前导序列，Ts 转录间隔序列，上图涂阴影区为转录单位，未涂阴影区为非转录间隔区

由于滑爪蟾二倍体细胞中有两个 rRNA 基因的同源区，因此可以认为双线期（DNA 已复制）的卵母细胞含有 4×450 约 1 800 个 rRNA 基因操纵元。但这还不够，还不能满足卵子形成的需要。根据对卵母细胞和体细胞中 28SrRNA 基因的比较分析，这一基因可能在卵母细胞中扩增了 15 000 倍。蛙类的体细胞都有 2 个与染色体相连接的核仁，但在卵母细胞中都有数千个核仁，而且都不与染色体相连，表明这些扩增的基因都是以染色体外因子存在。如果每一核仁都含有 450 个 rRNA 基因操纵元，那么在两栖类动物的卵母细胞中大约有 6.8×10^5 个 rRNA 基因的转录单位。

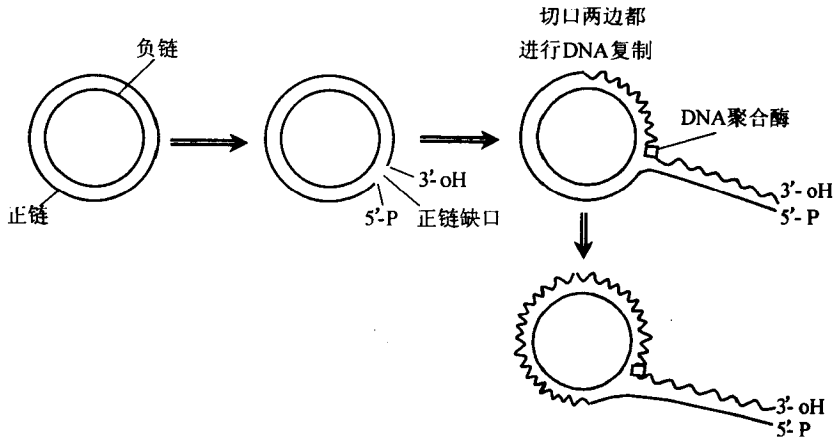


图 4－8　rRNA 基因扩增的滚环模型

通过复制产生许多相同的 rRNA 基因单位；环含有一个 rRNA 转录单位和一个非转录间隔区

有关 rRNA 基因扩增的过程目前还不清楚，但一般认为，在卵子发生过程中，卵母细胞基因组的一个 rRNA 基因转录单位首先通过某种机制脱离染色体，形成环状结构，然后这种染色体外因子按滚环模型（rolling circle model）进行复制（图 4－8）。复制开始时，环状 DNA 首先在某一位点形成缺口，然后再从缺口开始进行 DNA 合成，形成许多相同的 rRNA 基因转录单位。

（二）果蝇卵壳蛋白基因扩增

在普通果蝇的卵泡细胞中有 2 个卵壳蛋白基因簇，其中一个位于染色体 3 上，含有 4 个首尾连接的基因；另一基因簇位于 X 染色体上，含有 6 个卵壳蛋白基因。卵壳蛋白由卵泡细胞（ovarian follicle cell）合成，在卵壳蛋白基因表达之前，卵泡细胞的整个基因组首先进行多轮 DNA 合成，然后又有选择地复制卵壳蛋白基因。卵壳蛋白基因扩增时，通常不脱离染色体，只是在染色体的某些区域形成分叉（图 4－9）。每一分叉都含有若干扩增的卵壳蛋白基因，因此基因扩增以后，卵壳蛋白基因仍与染色体相连。卵壳蛋白基因扩增由叫做扩增控制因子（amplification control element）的顺作用子控制，扩增控制因子中可能含有特殊的复制启始区和促进 DNA 进行额外的、多轮次复制的调节序列。

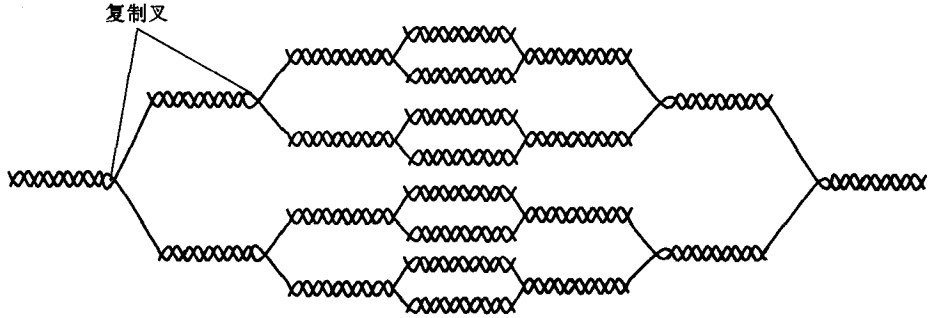


图 4－9　果蝇卵壳蛋白基因扩增模型

在染色体的某个含有多拷贝的卵壳蛋白基因区，该基因进行多轮 DNA 合成

（三）环境压与基因扩增

外部环境压也能使某些特殊基因进行扩增并使扩增的额外拷贝在基因组中固定下来。例如，培养的原仓鼠（chinese hamster）细胞用氨甲蝶呤（methotrexate）处理时，大多数细胞都死亡，只有少数细胞能成活。氨甲蝶呤是二氢叶酸还原酶（dihydrofolate reductase）的一种抑制剂，这些存活的细胞都能大量合成二氢叶酸还原酶，因此细胞含有足量未被抑制的还原酶来催化四氢叶酸（tetrahydrofolate）的合成，使细胞得以生存和生长。这些存活下来的细胞变异体能够将单拷贝的二氢叶酸还原酶基因扩增到 100～1 000 倍。

扩增的二氢叶酸还原酶基因既可以整合到染色体上，也可以作为染色体外

因子存在。在有丝分裂的中期，这种染色体外因子成对存在，叫做双微染色体（double minute chromosome）。双微染色体无着丝粒，当培养基中不含选择药物氨甲蝶呤时，它们又很快丢失。相反，扩增的二氢叶酸还原酶基因整合到染色体上后，常以一种延长的染色体带出现，这种染色体带叫做均一着色区（homogeneously staining region）。因此，这种扩增的基因是能够稳定遗传的。

无论扩增的二氢叶酸还原酶基因是位于正常染色体上，还是位于双微染色体上，它周围的DNA序列都与原来的未扩增的基因一样，保持不变，所以其调节表达和转录都是正常的。但是，扩增单位的大小却变化很大，有时只有25kb，只可包容二氢叶酸还原酶基因自身，有时可达到2 000kb。目前对这种基因扩增的机制还不清楚，可能基因组中有一些特殊的DNA序列与基因扩增有关。

第三节　基因的转录与蛋白质合成

如以前各章所述，转录是基因表达的第一个阶段，在转录之前，转录酶必须位于基因组的、待转录基因上游某个特殊位点上，将双链DNA的两条单链分开，启始转录过程并使RNA分子按基因组的遗传信息延长，最后在基因组的特定位点上终止。RNA在转录以后还需通过进一步加工才能形成成熟的RNA分子。

一、转录酶

真核生物细胞核基因的转录涉及一组依赖于DNA的RNA聚合酶，根据其亚基结构，所转录基因的类型以及对α-鹅膏蕈碱的抗性差异，将其分成三类即RNA聚合酶Ⅰ、Ⅱ、Ⅲ（表4-2）。这三种转录酶均为多亚基蛋白，在有模板DNA、三磷酸核苷、镁离子或锰离子的条件下具有转录活性。

RNA聚合酶Ⅰ存在于核仁之中，在不同生物中，其变化很大。例如在酵母中，该酶有12个亚基；而在小麦和花椰菜种质细胞中，该酶分别为7个和6个亚基。在酵母中，这种酶的转录产物为一条35S的前体RNA链，其中包括18S、5.8S和25S的rRNA以及长7kb的前导序列。

RNA聚合酶Ⅱ存在于核质当中，由8～14个亚基组成，分子量变动于14kd～220kd之间，主要功能是转录编码蛋白质的基因，对α-鹅膏蕈碱极为敏感。

RNA聚合酶Ⅲ也由8～14个亚基组成，分子量为16kd～155kd，主要功能是转录低分子量的序列，包括5SrRNA、所有tRNA，某些snRNA（也有某

些 snRNA 由 RNA 聚合酶Ⅱ转录）以及某些小分子 RNA 如 RNA 蛋白质复合体（SRP）中的 7SLRNA 等。此外，某些真核生物病毒的基因也由 RNA 聚合酶Ⅲ转录，如腺病毒的 VA RNA、EB 病毒的两种 EBER RNA。RNA 聚合酶Ⅲ对 α-鹅膏蕈碱不很敏感，高浓度的 α－鹅膏蕈碱才能抑制其活性。

表 4－2　3 种细胞核 RNA 聚合酶的特征

	聚合酶Ⅰ	聚合酶Ⅱ	聚合酶Ⅲ
存在部位	核仁内	核质内	核质内
转录产物	rRNA	前体 RNA	5SrRNA
		(hnRNA)	tRNA
亚型	2 种	2～3 种	1～2 种
α－鹅膏蕈碱	抗	极敏感	中度敏感
亚基数			
酵母	12	9	13
家鼠	6	9	10
小麦种质	7	10	—
滑爪蟾	—	—	14

二、转录的控制信号

(一) 真核生物蛋白质基因的启动子区

在原核生物中，基因的转录启始过程由叫做启动子、操作子的一些叫做顺作用子的特殊 DNA 序列控制，它们位于基因转录区的 5’端，是细菌基因表达必不可少的结构。在真核生物基因中，能够准确而有效地启始转录过程的 DNA 序列统称为启动子区。对由 RNA 聚合酶Ⅱ转录的、编码 mRNA 的基因（又称类型Ⅱ基因，class Ⅱ gene）来说，启动子区可分成几个部分（图 4－10）即帽子位点（cap site）、TATA 框（又叫 Goldberg－Hongnessbox）、CAAT 框和增强子等。由于增强子可以位于转录起点上游相当远的位置，也可以位于转录起点下游某个位置，所以对于其是否属于启动子范畴尚看法不一。

1.TATA 框和帽子位点　由 RNA 聚合酶Ⅱ识别的基因的控制序列差别非常大，其启动子中保守程度最高的结构为 TATA 框，即一段富含 AT 碱基的 DNA 序列，其共有序列为 5’ $TATA^{A}_{T}A^{A}_{T}$3’。在大多数真核生物的蛋白质基因中，这一位点距离转录起点大约 30bp；在少数基因中则不存在 TATA 框，可能其功能由某种相似结构所取代。TATA 框是一种定位因子，它控制转录机器

在其下游约 30bp 的位置启始 RNA 合成。

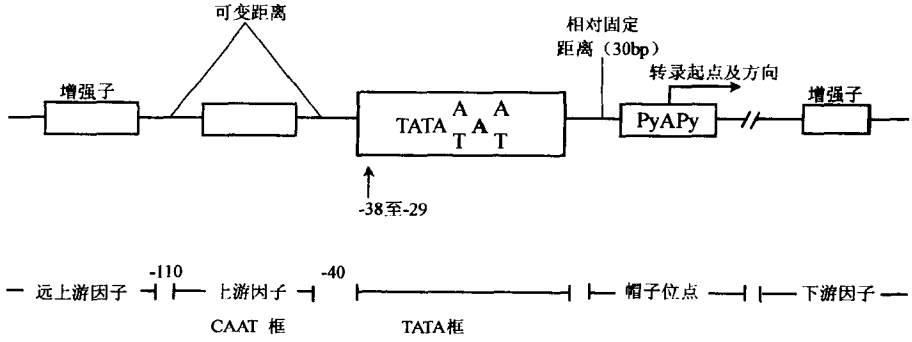


图 4－10　真核生物 RNA 聚合酶Ⅱ的启动子

第二个位点为帽子位点，它是转录起点所在位置，含有 mRNA5’端核苷酸，比 TATA 框保守性低。在人体血红蛋白基因中，帽子位点的序列为 ACATTTG。除掉帽子位点往往导致 RNA 转录起点发生改变和降低 RNA 聚合酶Ⅱ启始 RNA 链合成的速率。在某些情况下帽子位点还可能帮助转录机器启始 RNA 合成。

根据对基因离体表达的研究，细胞中至少存在一种与 TATA 框和帽子位点进行特殊互作的调节因子，三者必须首先形成稳定的前转录复合体，然后其他一些转录因子和 RNA 聚合酶才能启始 RNA 合成。

2. *上游启动子成员*　在 TATA 框上游－40～－110 的区域中存在几种结构保守序列，它们确定转录启始效率，但不决定转录启始的准确性。这些上游启动子成员具有多样性，其数目和离 mRNA 转录起点的距离因启动子不同而有差别，其结构保守性程度也不如 TATA 框高。

在某些动物和植物基因的启动子中，在 mRNA 转录位点上游－70～－110bp 位置，存在另一种结构保守序列，其同感序列为 GGPyCAA$^{T}_{A}$CT（又称 CAAT 框），如果这一序列中发生突变，将会使基因的转录水平降低。第二种同感序列为 CCPuCCC，它也是一种重要的转录控制序列，主要见于 SV_{40}的早期基因、疱疹病毒的胸苷激酶（thymidine kinase）基因和糖蛋白－D 基因以及 α－1 和 β－球蛋白基因。

另一类上游启动子成员只限于一些共同表达的基因家族中，如免疫球蛋白基因，热激蛋白基因、组蛋白基因和干扰素基因等。在免疫球蛋白基因家族中，轻链基因的可变区启动子含有结构保守序列 ATTTGCAT，而重链基因的启动子则含有与上述序列互补的结构 ATGCAAAT。在淋巴细胞中，这类结构

保守序列对 k－轻链基因表达具有重要作用，但它在另一些细胞系中可能没有什么功能。在果蝇中，热激蛋白基因也都具有一种相似的上游启动子成员 CTNGAANTTCNAG，至少其中有一部分序列对热激效应调节转录过程具有重要作用。

（二）增强子和沉默子

在真核细胞及其病毒基因组中，相对说来，增强子没有固定位置。自然状况下，它可以位于启动子上游如 SV_{40} 的早期基因增强子，也可以位于某个转录单位之内或其下游，如在免疫球蛋白基因中，重链和 k－轻链基因的增强子位于可变区和恒定区之间的内含子之中。在小牛乳头状瘤病毒的早期基因中，增强子则位于转录单位下游。另外，还有其他一些基因在转录起点下游具有提高基因转录水平的调节序列，可能这类结构也属于增强子范畴。

增强子具有组织特异性。例如免疫球蛋白重链基因的增强子通常只在 B 淋巴细胞内具有增强功能，如果将具有免疫球蛋白基因增强子的基因转入其他组织的核基因组中，该基因则不能转录。但是，如果将重链编码区的增强子插入到骨髓瘤（myeloma）细胞的血红蛋白基因之中，则可使该基因的转录水平提高 100 倍以上。

另一方面，基因组中那些与蛋白质激素结合有关的位点也可能具有增强子的功能。例如小鼠的乳腺瘤病毒中存在一组与糖皮质激素复合体结合的 DNA 序列，邻近这些位点的基因通常不活动，只有受到激素活化后才能进行转录。如果将一个不受激素调节的基因置于这类位点附近，该基因则能对糖皮质激素作出反应而进行转录。这类增强子含有结构保守序列 TGTTCT，它可以位于启动子前后约 500bp 的位置。这种现象说明，DNA 分子中某些区域通过与激素受体蛋白结合，以某种方式活化启动子而启始转录过程。

有关增强子的作用机理，目前存在几种不同观点。第一种观点认为，增强子可能是转录机器的入门位点，转录启始时，增强子可能形成一种特殊的 DNA 构型或者在染色体上某一位点（如 TATA 盖合）诱导某种构型变化，以促进某些转录因子之间的互作。第二种观点认为，增强子可能将 DNA 分子中待转录的序列引导到细胞核的某些特定区域如核基质中，以促进转录因子与启动子互作，因为这些区域富含转录因子。第三，增强子可能使染色质某一区域的结构发生变化，以促进某些转录因子与这一区域的 DNA 接触。第四，增强子可能直接与某些转录因子进行特殊互作。上述有关增强子的作用机制的观点之间并不相互矛盾，可能增强子具有上述所有功能。

真核生物基因组中还存在另一类与增强子功能相反、从远距离抑制基因转录的结构，叫做沉默子（silencer）。酵母交配系统（MAT）基因表达就受沉默

子控制。酵母染色体 3 上有 3 个非常相似的座位，即 MAT、HML 和 HMR（第八章）。当 MAT 表达时，另外两个座位则不表达，这一过程受距离 MAT 约 1kb 的沉默子调节。沉默子可能与接受某种外部信号有关。在外部信号的作用下，某种沉默子蛋白（silencer protein）与沉默子结合，从而抑制周围基因表达。

（三）座位控制区

座位控制区（locus Control region，LCR）又叫显性控制区（dominant control region）或座位活化区（locus activating region），属于远上游控制序列之一。起着稳定开放染色质的作用，并使与其邻近的基因或基因簇不依赖于其在基因组中的所取位置如何，独立地进行转录。在人体 β－珠蛋白（β－globin）基因簇的染色质中，座位控制区位于该基因簇上游 10～20kb 的位置，其中含有 4 段对核酸酶敏感的区域。如果将座位控制区与一个人体 β－珠蛋白基因（包括启动子和其他调节序列）偶联，然后再将其引入转基因小鼠中，结果人体 β－珠蛋白 mRNA 的合成量与整合的基因数目成正比，表明座位控制区使该基因独立于所在位置的染色质环境而进行表达。如果无座位控制区，则整合的 β－珠蛋白基因的表达水平很低，且与基因的拷贝数无相关性。虽然座位控制区的功能类似增强子，但在大多数情况下又不同于增强子规范的活性，因为座位控制区的自身的活性也依赖于染色质的结构。

座位控制区中含有多段与反作用子结合的序列，所以座位控制区的基本功能是调节协调互作过程，使启动子不含组蛋白，从而保证与启动子结合的反作用子在 DNA 复制期间停留在原位，形成敏感启动子。另外，座位控制区中不同部分也可能与不同启动子互作而表现出增强作用。

（四）RNA 聚合酶Ⅲ的识别位点

由 RNA 聚合酶Ⅲ识别的启动子可以分成 2 类，二者分别叫做内部启动子（internal promotor）和上游启动子（upstream promotor），由不同转录因子以不同方式加以识别。5SrRNA 基因和各种 tRNA 基因的启动子位于转录起点下游、转录单位之内。某些 snRNA 基因的启动子位于转录起点上游，与蛋白质基因的启动子相似。这两种启动子中都存在一些不同的与启动子功能有关的、由各种转录因子识别的启动子成员。

对内部启动子的研究最早见于滑爪蟾。例如，将滑爪蟾的 5SrRNA 基因的 5’和 3’侧翼序列除掉以后，该基因仍然能够准确转录，因为其 RNA 聚合酶Ⅲ的识别位点在 +50～+83 的一段序列之内。在蛾类（*Bombyx mori*）的酪氨酸（tyrosine）tRNA 基因和腺病毒 VA－1 基因中都存在类似结构保守序列（图 4－11）。这种控制信号含有 2 个结构保守区，分别叫做 A 框和 B 框，但这

两段序列之间的距离因物种不同和基因不同变化很大。无论在高等真核生物还是在原核生物中，tRNA 基因的这两段序列都是保守的。

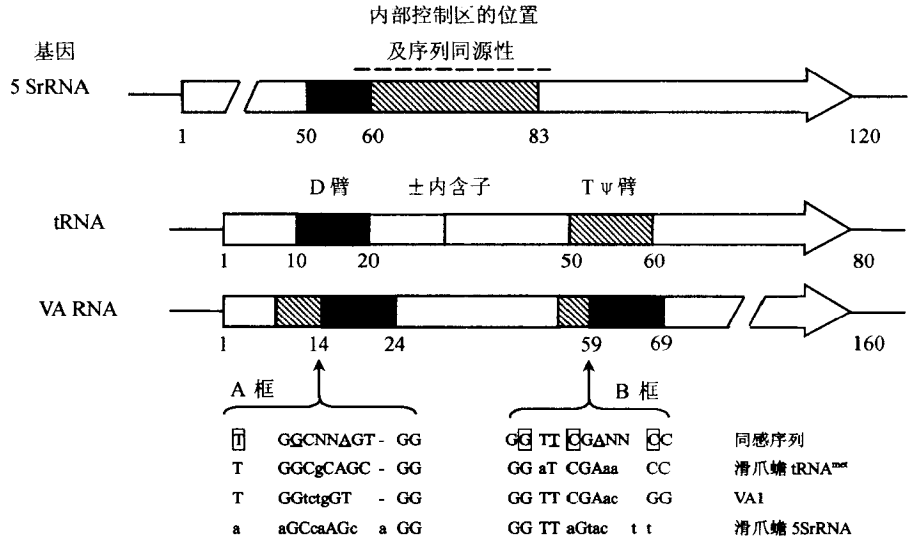


图 4－11　由 RNA 聚合酶Ⅲ转录的基因的内部控制区

数字从成熟 RNA 的 5’端开始。在同感序列的 A 和 B 框中，方框内碱基为高度保守的碱基（除 5SrRNA 基因外），加横线的碱基为所有 tRNA 基因中不变的碱基

除了内部控制区以外，RNA 聚合酶Ⅲ在转录过程中还需要基因上游或下游的侧翼序列。这些外部区域和内部信号都是 DNA 结合蛋白的识别位点，而这些蛋白质都是辅助转录启始因子，它们通过与 RNA 聚合酶Ⅲ互作，启始转录过程。上游启动子由三部分组成，即 TATA 框、近侧序列元件或称 PSE 因子（proximal sequence element）和 oct 因子（即一段 8 核苷酸序列）。这些启动子成员的作用与聚合酶Ⅱ启动子中相应成员的作用相似，其中 TATA 框可能决定聚合酶Ⅲ的特异性。

（五）RNA 聚合酶Ⅰ的识别信号

RNA 聚合酶Ⅰ只合成一种产物，即前体 rRNA 转录子。根据对不同物种 rRNA 基因转录起点的研究，其转录起点周围的序列变化很大，只是在非常相近的物种之间，这一区域的序列才存在同源性。根据对哺乳动物 rDNA 单个碱基突变的研究，在转录起点上游－30 位存在三个重要碱基 TTT，在－16 位存在高度保守的 G。另外，可能转录起点下游的几个碱基也对 RNA 聚合酶Ⅰ准确而有效地启始转录过程具有重要作用。由于 rRNA 转录起点周围的这段序列对 RNA 聚合酶Ⅰ启始转录具有重要功能，因此，通常认为这组序列是 rRNA 基因的启动子。

在滑爪蟾中，各个 rRNA 操纵元的转录也受两个相邻操纵元之间非转录间隔序列的影响。每一间隔序列都由若干 DNA 重复序列组成，其长度不等，其中有一段长 200bp 的重复序列与 rRNA 操纵元的启动子存在 90% 的同源性，因此这段序列也具有启动子功能。这种重叠启动子与真正启动子之间还存在其他一些长度为 81 到 60bp 的重复单位（图 4－12）。由于这种 60/81bp 的重复序列具有提高其下游 rRNA 基因启动子启动转录的作用，因此它们可能是某些 rRNA 转录调节蛋白的结合位点。

根据对人体 DNA 的研究，RNA 聚合酶Ⅰ的启动子由两部分组成，其中之一叫做核心启动子（core promotor），即位于转录起点周围从－45～＋20 的一段序列，单独能启始转录过程。第二个区域叫做上游控制因子（upstream control element UCE），为－180～－107 的一段序列，均能极显著地提高核心启动子的转录启始效率。二者都富含 G－C 碱基对，而且约有 85% 的序列同源性。

rRNA 操纵元的转录具有明显的物种特异性。可能转录启始因子之间及其识别序列之间的差别是造成这种物种特异性的基础。例如，小鼠细胞抽提物只转录小鼠 rDNA，而人体细胞抽提物只转录人体 rDNA。但是，当将这两种细胞抽提物分离以后，发现 RNA 聚合酶Ⅰ并不具有物种特异性。造成 rDNA 转录的物种特异性的则是其它额外因子。在这些额外因子中至少有一种因子与 RNA 聚合酶Ⅰ互作，结合到转录起点附近的 DNA 序列上，形成稳定的转录复合体。另一方面，在人体和小鼠细胞融合后形成的杂种细胞中，通常人体染色体逐渐丢失，杂种细胞只含有一个亲本的染色体。虽然两个物种的 rRNA 基因都能被保存下来，但杂种细胞只能表达保留了染色体的那一个亲本 rRNA 基因，这可能是亲本之一的染色体丢失后，位于其上的、编码物种特异性 rRNA 转录因子的基因也随之丢失的结果。

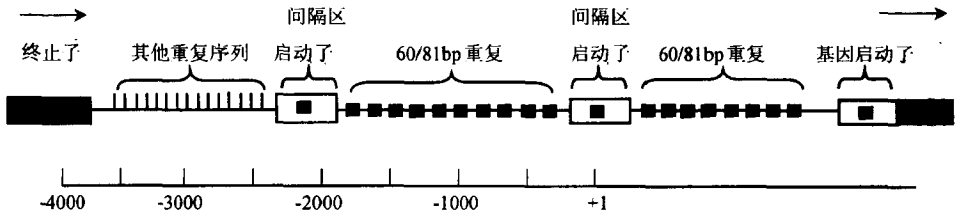


图 4－12　滑爪蟾 rDNA 的一个间隔非转录区

其大部分序列都含有与真正启动子有关的序列，小黑块为与转录起点上游－72～－114 之间的 42bp 序列存在高度同源性的序列，未涂黑区（其中含有黑块）为 200bp 的不完全重复，间隔区的一端即接近转录终止子的区域为与启始转录无关的重复序列（根据 R.H.Reader，1984）

（六）转录终止信号

1.RNA 聚合酶Ⅱ的转录终止信号　有关真核生物中蛋白质基因的转录终

止信号目前还不清楚。根据对动、植物组蛋白基因的研究，几乎所有组蛋白基因的3'端都有一个双向对称结构，这种结构可能是组蛋白基因的转录终止信号。例如在海胆的组蛋白基因中，这种发夹结构中的碱基配对部分由（G）GCPyCT（T）和AGPuC（C）组成，其环状部分通常为TTN。又如，小麦的H_4基因3'端也有一个双向对称结构，未配对的环中的碱基也为TTN，与多细胞动物发夹环中的碱基组成非常相似。

在原核生物中，比较典型的转录终止信号为一串碱基T，虽然大多数高等真核生物的基因不具有这种结构，但是某些基因如核糖体蛋白基因的终止密码下游都有5～6个连续的碱基T，可能这种T碱基串具有终止转录的功能。

2.RNA聚合酶Ⅲ的转录终止信号　聚合酶Ⅲ的转录终止信号同其启动子一样，也位于转录区之内，一般在其非模板链上都有4个以上T碱基串，而且在这一区域前后还接有富含碱基GC的序列，这种结构具有原核生物的典型转录终止信号的特征。

3.RNA聚合酶Ⅰ的转录终止信号　在脊椎动物中，rRNA操纵元的转录终止信号位于28S序列末端以外数百个碱基的位置，目前尚不清楚RNA聚合酶Ⅰ是否能普遍性地识别这种信号。

三、转录启始及转录因子

（一）转录机器与转录活化

1.RNA聚合酶Ⅱ的转录启始复合体　编码蛋白质的基因都由RNA聚合酶Ⅱ转录，但这种转录酶单独不能转录基因，它还需要许多额外蛋白质的帮助才能识别启动子和启始转录。这些指导转录的蛋白质（包括聚合酶Ⅱ）组成一个小复合体（图4－13），叫做全酶，一起行使功能。这类辅助蛋白中有一

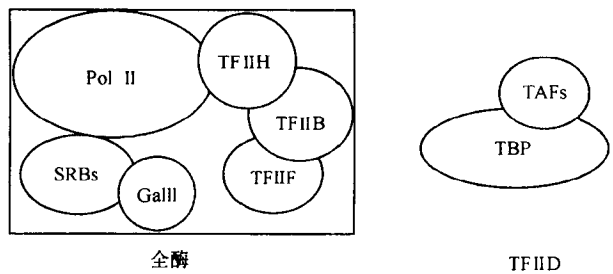


图4－13　酵母的转录启始机器
全酶中各种成分都由多蛋白质组成

些为普通转录因子（transcription factor，TF），包括 TFⅡA、TFⅡB、TFⅡD、TFⅡE、TFⅡF、TFⅡH 等。其中有些因子含有一种以上成分，如转录因子 TFⅡD 含有一种与 TATA 序列结合的蛋白质（TATA binding protein，TBP）和一组叫做 TAF 的蛋白质（即一类与 TBP 结合的因子）。在酵母和哺乳动物中，活体内转录除需要各种 TF 外，还需要一组叫做 RNA 聚合酶Ⅱ抑制子或称 SRB（supressor of RNA polymerase B，B 为 RNA 聚合酶Ⅱ）的调节蛋白。

TBP 是最先与聚合酶Ⅱ启动子结合的转录因子，由于 TATA 框距转录起点的距离是相对固定的，所以对 TATA 框的识别对定位 RNA 聚合酶是非常重要的。

TBP 只与 DNA 分子的小沟结合，而几乎所有已知的 DNA 结合蛋白都与大沟结合。TBP 具有晶体状结构，沿 DNA 双螺旋的一个面结合，形成一种马鞍状结构。马鞍结构的内表面与 DNA 结合，外表面则与其他蛋白质互作。TBP 的 DNA 结合位点的序列在不同物种之间是保守的，暴露在外的氨基末端部分则与其他蛋白质互作。TBP 以一种特殊方式与 DNA 结合，它不仅只结合到 DNA 小沟之中，而且还使 DNA 弯曲 80 度。TATA 框朝主沟弯曲，使小沟变宽。这种变形只局限于 8bp 的 TATA 框。

当 TBP 与启动子结合后，TFⅡD 就成为其他与启动子结合的转录因子集中的中心。这些转录因子与启动子结合的顺序依次为 D、B、F、聚合酶Ⅱ、E 和 H。TFⅡA 可以在任何时间与启动子结合。RNA 聚合酶Ⅱ和 TFⅡF 同时与启动子结合，TFⅡF 将聚合酶Ⅱ补给到启始复合体上。结合了各种转录因子和 RNA 聚合酶Ⅱ的启动子叫做前启始复合体（preinitiation complex）（图 4－14）。

对于缺少 TATA 框的基因来说，含有 TBP 的 TFⅡD 仍然是这类基因转录的重要成分，同样行使组装功能，但它需要一些额外因子以帮助其与 DNA 结合。

除了这些普通转录因子外，真核细胞中还存在一些基因特异性的转录因子，其中有一些与 GC 框和 CCAAT 框结合，还有一些则与更特殊的 DNA 序列如增效子等结合。

2.RNA 聚合酶Ⅰ的转录启始复合体　现在知道哺乳动物的 RNA 聚合酶Ⅰ需要两种转录辅助因子才能转录 rRNA 操纵元。一种叫上游结合因子（upstream binding factor，UBF），以一种序列特异性方式结合到核心启动子和上游控制成员的相关序列上。另一种转录辅助因子叫做 SL_1，这种辅助因子本身不具有启动子特异性，但是一旦 UBF 与启动子结合后，SL_1 就能与启动子结合并与 UBF 进行互作。当这两种转录辅助因子结合以后，RNA 聚合酶Ⅰ就与

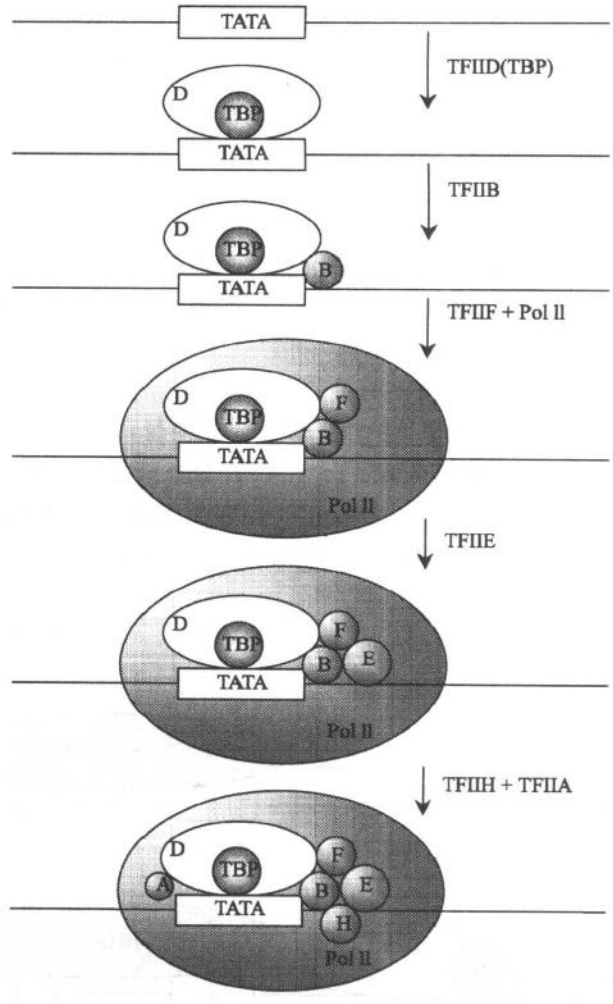


图 4－14　RNA 聚合酶Ⅱ前转录启始复合体的组装

TFⅡA 在组装过程中可能较早与启动子结合

核心启动子结合并启始转录（图 4－15）。

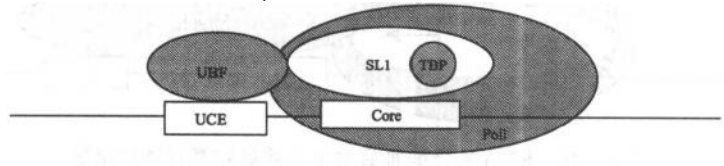


图 4－15　RNA 聚合酶Ⅰ的前启始复合体

UCE 和 Core 分别为启动子的两部分

UBF 只有一条多肽链，与核心启动子和上游控制成员中富含 G－C 碱基对的一段序列结合。UBF 和 RNA 聚合酶Ⅰ都不具有物种特异性，如家鼠的 UBF 和聚合酶Ⅰ能识别人体的 rRNA 操纵元。而 SL_1 则具有物种特异性，如家鼠的 SL_1 不能启始人体 rRNA 基因转录，人体的 SL_1 则又不能启始家鼠 rRNA 基因转录。

SL_1 由 4 种蛋白质组成，TBP 是其成分之一。由于 TBP 不具物种特异性，可能 SL_1 的物种特异性和 DNA 结合功能由其他成分所提供。SL_1 的作用类似细菌的 σ 因子，不能与启动子特异结合，而是通过与其他成分互作，然后再结合到启动子的特定区域。

3. RNA 聚合酶Ⅲ的转录启始复合体　RNA 聚合酶Ⅲ行使功能时涉及三种转录因子，即 TFⅢA、TFⅢB 和 TFⅢC，这些蛋白质先于 RNA 聚合酶Ⅲ与启动子结合，形成前启始复合体，指导 RNA 聚合酶与启动子结合。TFⅢA 是锌指蛋白（zinc finger protein）之一，TFⅢB 由 TBP 和另外两种蛋白质组成，TFⅢC 为一种大蛋白质复合体（500kd 以上），至少含有 5 个亚基。一般认为

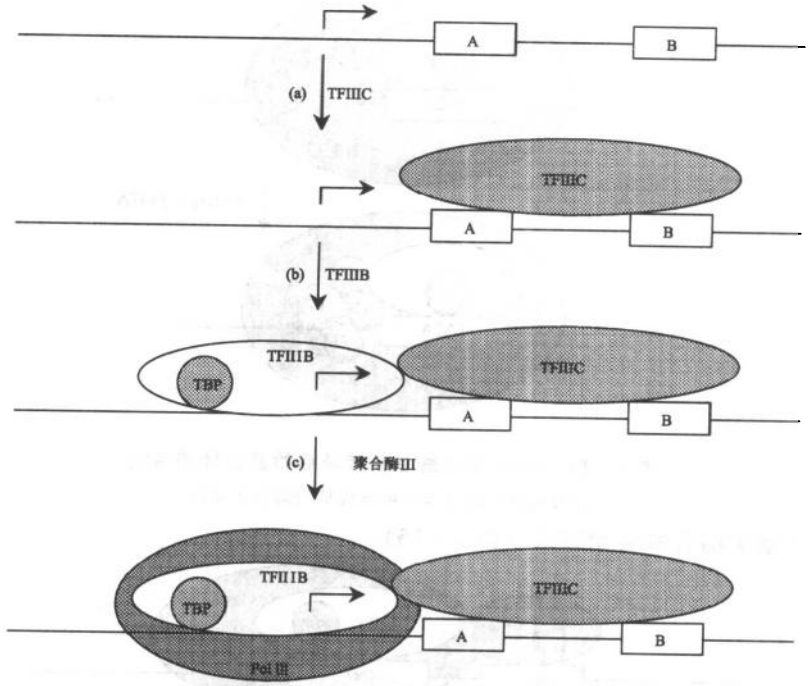


图 4－16　RNA 聚合酶Ⅲ启动子上前启始复合体的组装

A. TFⅢC 与内部启动子的 A 和 B 框结合　B. TFⅢC 促进 TFⅢB 以其 TBP 与转录起点上游的区域结合　C. TFⅢB 促进聚合酶Ⅲ结合到转录起点上，开始转录

TFⅢB是RNA聚合酶Ⅲ所需的转录启始因子，而TFⅢA和TFⅢC则为组装因子，其作用是帮助TFⅢB结合到正确位置上。TFⅢB是RNA聚合酶Ⅲ的定位因子，使聚合酶Ⅲ准确结合到转录起点上（图4－16）。TFⅢB也同细菌RNA聚合酶中σ因子相似，本身不能与DNA直接结合，但同其他蛋白质一起则能结合到DNA上。

TBP是RNA聚合酶Ⅲ转录各类基因所必需的。对于具有TATA框的启动子而言，TBP能与DNA进行特异结合，但对于其他启动子而言，TBP则通过与其他与DNA结合的蛋白质结合而被吸收到转录启始复合体上。作为TFⅢB的成分之一，在使用内部启动子启始转录期间，TBP与RNA聚合酶Ⅲ的定位有关。作为SL_1的成分之一，TBP则将RNA聚合酶Ⅰ定位到rDNA上。作为RNA聚合酶Ⅱ的相关启始因子，它直接识别TATA框。因此，TBP是各种RNA聚合酶的关键定位因子，通过不同机制被吸收到各类启动子上。无论TBP以何种方式进入转录启始复合体，它只有一个目的，即与RNA聚合酶进行互作。

（二）RNA聚合酶Ⅱ的基因特异性转录因子

1. *与启动子成员互作的转录因子*　除了普通转录因子以外，真核细胞中还存在各种不同的基因特异性转录因子。其中有些与启动子中CCAAT框或GC框结合，还有一些则与某些基因的增强子互作。在人体的肿瘤细胞（Hela细胞）中有一种叫做Sp_1的蛋白质（Sp_1Protein），这种转录因子可以极大地促进由SV_{40}早期启动子控制的转录和由HSV胸苷激酶基因启动子控制的转录。与此相反，Sp_1能够降低另一种人体病毒腺病毒（adenovirus）的、由主后期启动子控制的转录水平。

Sp_1的重要功能是与某些启动子的5'区中GC框结合，促进由这类启动子控制的转录。例如SV40的早期启动子中在TATA框和增强子之间有一段21bp的GC重复，每一GC框之间的距离为DNA双螺旋旋转两周，使结合后的Sp_1分子都处于DNA分子的同一面上（图4－17）。

在HSV病毒的胸苷激酶启动子及其他启动子中，CCAAT框则与其独特的转录因子互作，这种转录因子叫做CCAAT框转录因子（CCAAT box transcription factor，CTF）（图4－17）。

2. *与增强子互作的转录因子*　与Sp_1和CTF转录因子相比，与增强子结合的转录因子更为特殊，因为具有增效子的基因数目比具有GC框和CCAAT框的基因少得多。

在酵母中有一种控制一组与半乳糖代谢有关的基因表达的活化蛋白GAL4蛋白，这种蛋白质是基本的基因特异性转录因子或增强子结合蛋白之一。每一

受 GAL4 调节的基因在其转录起点上游都有一个 GAL4 结合位点。这种结合位

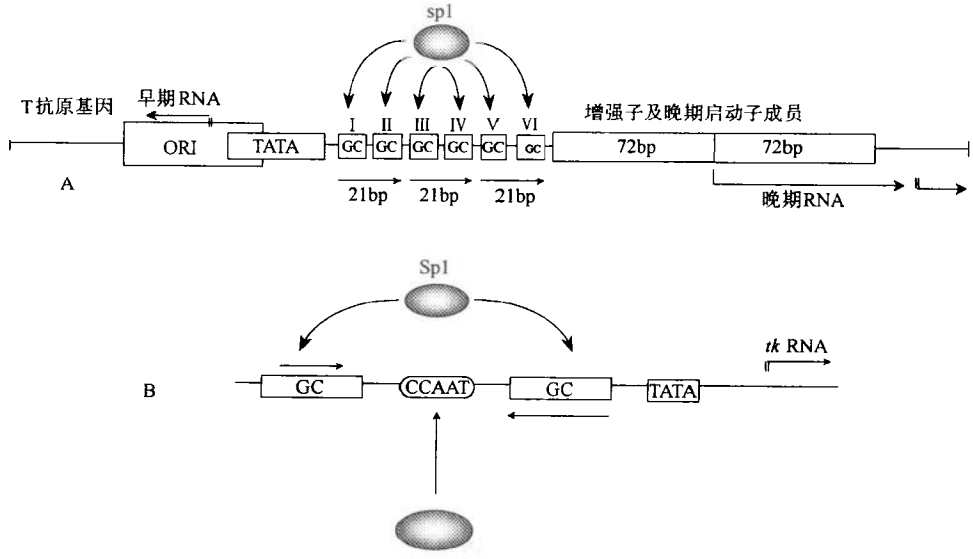


图 4－17　转录因子与两种启动子的互作模型

A. 转录因子 Sp_1 与早期启动子中 6 个 GC 框互作　B. 转录因子 Sp_1 和 CTF（CCAAT 框转录因子）与胸苷激酶启动子互作

点叫做上游活化序列（upstream activating sequence UAS）。GAL4 以二聚体形式与 UAS 结合，活化下游启动子的转录（见本章后）。

核受体（nuclear receptor）为另一类重要的与增强子结合的蛋白质。这类蛋白质能与各种类固醇及相关激素互作，形成激素/受体复合体，这种复合体通过与叫做激素反应子（hormone response element）的增强子结合，促进有关基因转录。这类激素主要是一些性激素、孕酮、糖皮质激素、盐皮质激素、维生素 D、甲状腺激素和视黄酸等。每一种激素都与其特殊受体结合，然后再活化各自控制的基因的表达。某些核受体如糖皮质激素受体存在于细胞质中，与热激蛋白 90（hsp90）形成复合体。当这种复合体在细胞质中遇到其相应的配位体或激素时，便脱离 hsp90 并与激素结合。然后这种激素／受体复合体再进入细胞核内，调节有关基因表达（图 4－18）。另一些受体如甲状腺激素受体只存在于细胞核中，在无激素时，它与有关增强子结合并抑制转录。如果存在激素，这类受体则形成激素/受体复合体，然后再结合到相应增强子上，行使活化子功能。

（三）基因特异性转录因子的结构

许多转录因子包括增强子结合蛋白都具有相似结构，都由三个功能部分组

成，即 DNA 结合域、转录活化域和二聚体化域。

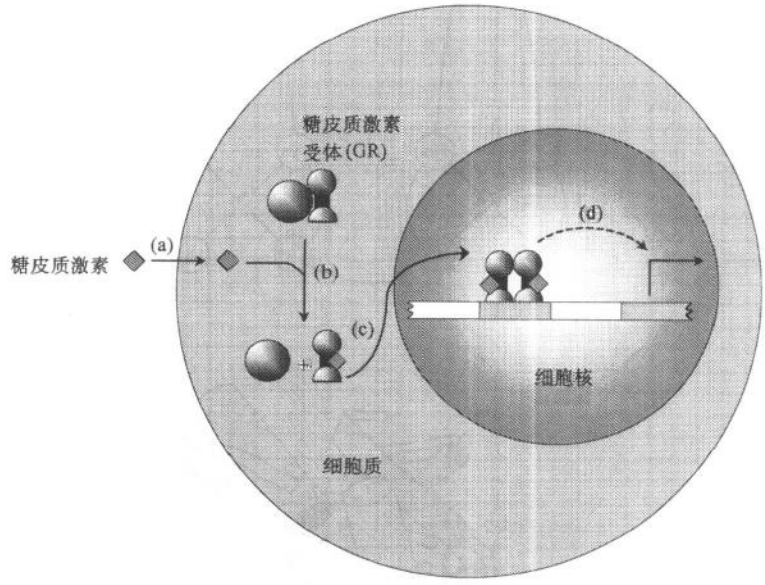


图 4－18　糖皮质激素的作用模型

a. 糖皮质激素通过扩散，从细胞外进入细胞质　b. 糖皮质激素与其受体（GR）结合并使受体脱离 hsp90　c. 激素/受体复合体（HR）进入细胞核并与另一 HR 形成二聚体，然后再结合到有关基因上游的激素反应子或增强子上　d. 结合到增强子上的 HR 二聚体活化有关基因转录

1. DNA 结合域　DNA 结合域（DNA－binding domain）又可分成三个亚区即锌指（zinc finger）、同源异型域（homeodomain HD）和亮氨酸拉链域（leucine zipper）。

锌指首先见于 RNA 聚合酶Ⅲ的转录因子 TFⅢA，后来发现其他许多转录因子都具有锌指。根据对转录因子 Sp_1 的结构研究表明，它具有一个指头状的重复区（图 4－19A），每一重复都与一个锌离子结合，所以就把这种结构叫做锌指。

在家鼠基因组中有一个所谓的立即早期基因（immediate early gene），该基因是静止期的哺乳动物细胞开始分裂时最先被活化的基因，编码一种叫做 Zif268 的锌指蛋白。Zif268 蛋白具有 3 个相邻的、填充到 DNA 双螺旋主沟中的锌指（图 4－19）。Zif268 蛋白的每一锌指都有一个 α－螺旋（圆柱状结构），它通过其底部的一个小环与指头的另一侧相连。Zif268 的 3 个指头的形状几乎完全相同，它们呈曲线型分布或呈 C 型，填充到 DNA 双螺旋的主沟中。所有指头都以基本相同的角度朝向 DNA，所以这种蛋白质/DNA 接触的

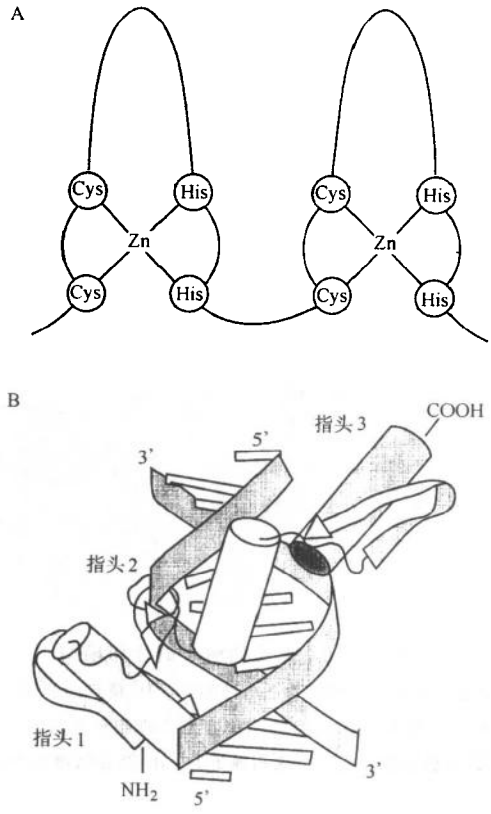


图 4－19　锌指的结构模型

A. 两个相邻锌指的结构模型　每一指头中的锌离子与 4 个氨基酸结合，2 个半胱氨酸和 2 个组氨酸使指头形成适合与 DNA 结合的构型。这 4 个关键氨基酸结合到同一指头的相同位置上　B. Zif268 的三个锌指形成的蛋白质/DNA 复合体的立体构型，锌离子未标出

几何图形非常相似。绝大多数锌指蛋白都通过其指头的 α－螺旋中的氨基酸同 DNA 主沟中的碱基接触，这是锌指蛋白的用于识别 DNA 的重要特征之一。类固醇激素受体也都有含锌部分，但不同的是与锌原子结合的为 4 个半胱氨酸（无组氨酸），所以，在形状上其含锌部分与典型的锌指有所不同。

同源异型域首先见于果蝇的调节发育的 DNA 结合蛋白中，后来发现在许多转录因子中都存在同源异型域。以果蝇的 OCT 蛋白（八聚体结合蛋白）为例，它有一段长 75 个氨基酸的、叫做 Pou 区的结构，这段区域与同源异型域相邻，二者组合在一起非常类似 λ 噬菌体抑制子的蛋白质结合区。所以，一般认为同源异型域与转录因子同 DNA 结合有关，含有对 DNA 进行特异性识别的结构。

大多数转录因子都以二聚体形式与 DNA 结合，而每一分子的转录因子通常都有一个叫做亮氨酸拉链的区域，正是这种亮氨酸拉链与转录因子的二聚体化有关（图 4－20）。

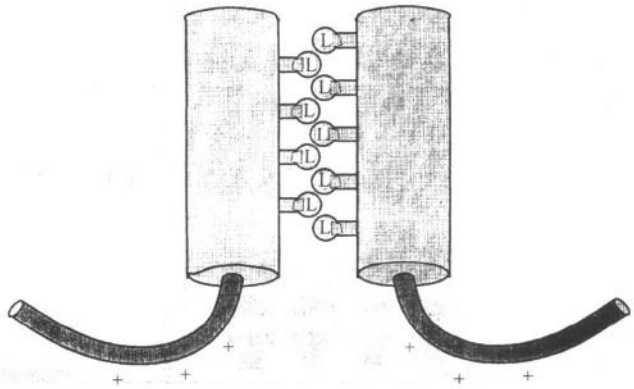


图 4－20　亮氨酸拉链模型

两条相同的蛋白质分子都有一个 α－螺旋（圆柱体），α－螺旋表面的亮氨酸以类似拉链的两部分的方式进行互作，当两条蛋白质形成二聚体后，亮氨酸拉链外的碱性区（带正电荷“＋”）与 DNA 结合

每一个分子的转录因子都有一个含有亮氨酸或其他疏水氨基酸的 α－螺旋。α－螺旋中每相隔 7 个氨基酸就有 1 个亮氨酸，从而使所有亮氨酸都位于 α－螺旋的外表，使其处于与另一蛋白质分子的相同部分的亮氨酸进行互作的位置。两个螺旋以这种方式进行互作就像一条拉链的两部分连系在一起一样（图 4－20）。亮氨酸拉链不仅将两条蛋白质分子连系在一起，而且还将其 2 个碱性区域填充到 DNA 双螺旋的主沟之中。

当然，这种模型并不能完全代表转录因子与 DNA 互作的方式，还有另外一些转录因子则以不同方式与 DNA 互作，这里不作详述。

2. *转录活化域*　大多数转录因子都有一个以上的活化域，这些活化域可以分成 3 类。第一类为酸性活化域，如酵母的转录因子 Gal4。Gal4 的活化域有 49 个氨基酸，其中就有 11 个为酸性氨基酸。第二种为富含谷氨酰胺的活化域。如转录因子 Sp_1 就有 2 个活化域，每一个约含 25% 的谷氨酰胺。此外，Sp_1 还有另外 2 个特殊的活化域。第三种为富含脯氨酸的活化域，如转录因子 CTF。

一般认为，活化域通过与其他转录因子如 TFⅡB、TFⅡD 或 RNA 聚合酶等进行互作，协助组装前转录启始复合体，从而促进转录。

（四）转录活化子的作用机制

大多数蛋白质基因的转录活化过程都受活化子调节。基因活化的补给模型

认为，活化子的作用是将转录机器补给到 DNA 上，但在真核细胞中，补给作用只发生在 RNA 聚合酶结合到启动子上之后。典型的活化子都有一个 DNA 结合部分和一个活化部分，其 DNA 结合部分与被调节基因上游的特殊位点如增强子等结合，其活化域则与转录机器中多蛋白质接触（图 4－21）。活化作用的特异性只由活化子的 DNA 结合位点所决定。例如，酵母中活化子 Gal4 通常只活化半乳糖代谢所需的基因，因为在这些基因上游都存在 Gal4 的结合位点。在离体条件下，活化子可与全酶的某些成分互作，也可与各种 TAF 互作。因此，可能活化子与转录启始装置的任何成分结合，都可活化活体内基因的转录。

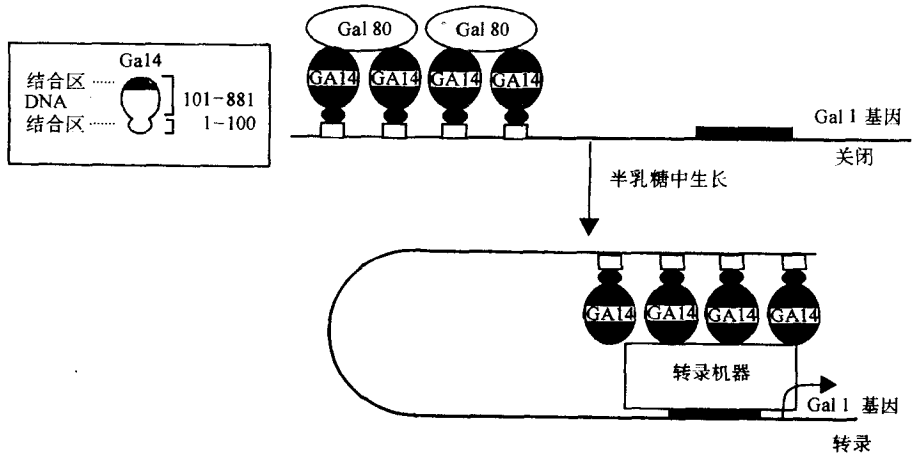


图 4－21　酵母活化子 Gal4 对半乳糖代谢基因的活化模型

Gal1 基因是半乳糖代谢所需基因之一；当培养基中无半乳糖时，Gal4 的活化域被抑制子 Gal80 遮盖；当存在半乳糖时，抑制作用被解除，Gal4 的活化域通过 DNA 环出与转录机器接触，启动活化作用；Gal1 基因上游有 4 个 Gal4 结合位点，一个 Gal4 二聚体可识别一个位点（根据 M.Ptashne 和 A.Gann，1997）

对于高等真核生物来说，许多基因的转录都依赖于多种生理信号，其有效转录涉及多种活化子－转录机器之间的互作，而且每一种互作都由各不相同的生理信号控制。因此各种活化子只有协同作用才能活化基因。

如前所述，真核生物基因的启动子都含有多个顺作用子，如哺乳动物的金属巯基组氨酸三甲内盐 I（metallothioneineI）基因的启动子除具有一个 TATA 框外，还有 1～2 个 Sp_1 蛋白结合位点，1～2 段重金属诱导该基因表达所需的 DNA 序列，一个与糖皮质激素（glucocorticoid）结合有关的区域和几个组成型增强子。同样，小鼠的具有组织专性表达特点的清蛋白基因的启动子也含有几段与序列特异蛋白结合的 DNA 序列，其中有三段序列与肝组织特异性因子结

合。其他一些启动子成员如 TATAAATA、CCAAT 和 GGGCGG 框都与不同反作用子结合。当然，在另一些基因中，可能启动子成员还包括光反应子、糖皮质激素反应子、热激反应子和重金属反应子等顺作用子。同原核生物基因一样，真核生物基因的顺作用子都位于转录起点附近，但某些真核生物基因的顺作用子，增强子或沉默子等可以从相当远的位置作用于基因的转录过程，它们可能通过与反作用子一起，与远端和近端的启动子成员结合，使二者之间的 DNA 环出而发挥其远距离作用，活化或抑制基因的转录过程。

基因表达的负控制有几种形式，如关闭某些不宜表达的基因或降低某些基因受诱导表达的水平，这种反作用子叫做转录降低调节子（transcriptional downregulator），它以几种不同方式行使功能。其中有些降低调节子只具有 DNA 结合部分，缺少功能活性部分，它们与活化子竞争同一 DNA 结合位点，阻止活化作用。另一种方式是，它们与活化子形成二聚体，降低活化子与 DNA 结合的亲和性或降低其活化转录的能力。此外，还有些抑制物在与 DNA 结合时，能与活化子互作，抑制活化子的活化功能。另一类转录降低调节因子包括一些抑制蛋白，它们将转录复合体中的活化因子隔开，使之不能与 DNA 结合。现在知道，不仅活化子和抑制子可以由不同基因编码，而且在某些情况下，同一个调节基因既可以编码活化子，也可以编码抑制子，而且活化子和抑制子在体内是协调合成的。

（五）影响转录的其他因素

1. *核骨架与基因转录启始*　相对原核细胞而言，真核细胞 DNA 与各种组蛋白结合，形成比较稳定的染色质。染色质在细胞核中并不自由浮动，间期染色质的某些区域都通过核片层（nuclear lamina）固定到细胞核内膜上。核片层为一种丝状蛋白质网络系统，位于核内膜表面，是核骨架（nuclear scaffold）中最明晰的结构。在细胞周期的各个阶段，某些特殊的 DNA 序列都牢固地结合到核骨架上。

根据对果蝇培养细胞的研究，其基因组 DNA 都在某个特殊位置上与核骨架相连。例如果蝇的每一个组蛋白基因重复单位（$H_1-H_3-H_4-H_2A-H_2B$）中都有一个与核骨架相连的附着点，这个位点是 H_1 和 H_3 基因之间的一段富含 AT 的非转录间隔序列。根据对果蝇其他基因的分析，这些基因附着到核骨架上的位点位于转录区上游、距转录区约 5kb 的一段序列之中，而且在两个附着点之间还可能存在一至数个基因。总的说来，积极表达的或将要表达的基因都通过一个与其相距很近的附着点与核骨架相连，而另一些不表达的基因则不与核骨架相连。

核骨架的功能是保留对转录具有重要作用的蛋白质，由于基因只是在启动

子附近附着到核骨架上，而不是在转录区内附着到核骨架上，我们可以想象，当 RNA 合成终止之后，沿游离 DNA 环滑动的 RNA 聚合酶就很容易被核骨架再次捕获，然后在适当转录因子的帮助下，再次在接近核骨架的位点上启始转录过程。

2. 核小体与基因转录启始　从理论上讲，不具有核小体的 DNA 都能被 RNA 聚合酶有效地进行转录，但转录并不一定需要消除核小体。例如在一种昆虫（*Oncopeltus fasciatus*）中，含有 rRNA 基因的 DNA 部分始终都存在核小体，所以这一区域的转录活性非常高。但在其他一些物种中，虽然 rRNA 基因都盘绕成核小体，但它仍然能够迅速转录。

基因转录活性与染色体中某些特殊的非组蛋白存在很高相关性，这类蛋白质可能与核小体结合，使有关基因处于转录准备阶段。例如，当染色质用 DNA 酶 Ⅰ 处理时，染色质降解后可以释放出两种非组蛋白即 HMG（high modility group）14 和 HMG17，如果首先将这两种蛋白质从染色质中除掉，染色质中相应的 DNA 区域就不能被 DNA 酶 Ⅰ 降解。这类蛋白质使染色质对 DNA 酶 Ⅰ 敏感的作用具有组织特异性，如从大脑中分离的 HMG14 和 HMG17 只能使红血细胞中的球蛋白基因对 DNA 酶 Ⅰ 敏感，而不能使脑细胞中的球蛋白基因对该酶敏感。因此，染色质中这两种非组蛋白可能是不同组织中某些基因进行转录所必需的辅助因子。

DNA 分子与组蛋白 8 聚体核心结合并不一定对转录有害。根据对离体条件下染色质模板转录过程的研究，一旦转录开始以后，RNA 聚合酶就能够转录含有组蛋白 8 聚体核心的基因，由 RNA 聚合酶Ⅱ和 Sp6 聚合酶催化的 RNA 链延长过程并不受一、两个组蛋白核心的阻碍。如果组蛋白 8 聚体较多，RNA 聚合酶Ⅱ仍然能够延长 RNA 链，但转录过程则部分受抑制。RNA 聚合酶能够转录核小体的机制目前还不清楚，它可能将 8 聚体或其他成分从转录复合体直接移到模板链的另一部分，使 8 聚体松散或直接破坏 8 聚体。

虽然 RNA 链延长并不明显地受核小体的阻碍，但是如果组蛋白 8 聚体与启动子结合，则转录启始会受到抑制，因此核小体通过阻隔某些关键 DNA 序列而明显影响转录过程。在转录启始之前，可能全酶及各种转录因子通过与组蛋白争夺启动子位点即所谓的动力竞争（dynamic competition），形成转录启始复合体。第二种可能性是各种转录因子在 DNA 复制期间就先于组蛋白与启动子结合，即所谓的前占竞争（preemptivecompetition），使启动子不含组蛋白（图 4－22）。显然，组蛋白是调节机器的一个重要部件，它阻碍全酶的有效补给。在形成转录启始复合体的过程中，可能增强子以及座位控制区等通过染色质弯曲与启动子结合而表现出增强作用。

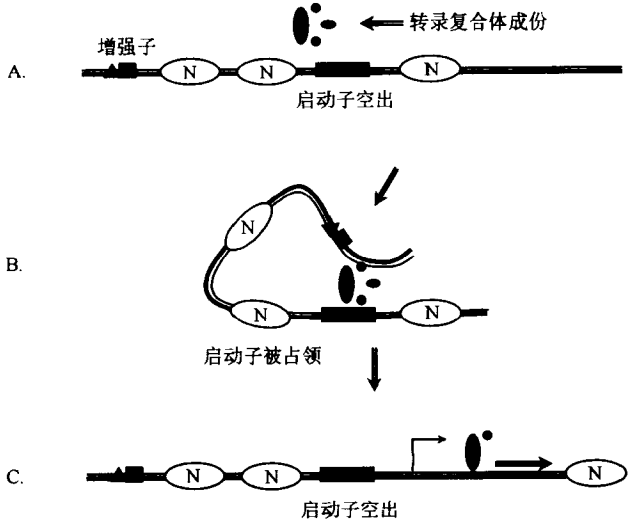


图 4－22　核小体与基因转录活化

A. 转录因子或全酶与启动子结合，使该区域不含组蛋白　B. 转录因子通过与增强子上其他蛋白质因子互作而被固定到启动子上，形成转录启始复合体　C. 转录开始后启动子空出 N－核小体

四、mRNA 的加工

真核生物的大多数基因都是隔裂基因，基因中编码蛋白质的序列（外显子）由若干非编码序列（内含子）隔开。根据对人体 β－珠蛋白基因转录子的研究知道，其前体转录子含有前导序列、外显子、内含子和 3' 非翻译区。这种初级转录子必须经过一系列加工步骤才能成为具有翻译功能的 mRNA。也就是说，细胞核中的加工机制必须将前体 mRNA 的内含子和某些多余的碱基除掉，将外显子连接起来，然后 mRNA 再进入细胞质中进行翻译。

（一）mRNA 帽子的形成

在所有真核生物中，成熟 mRNA 的 5' 端都有一个帽子结构（cap structure），它是一个 7－甲基鸟苷基团以一种极不寻常的 5' →5' 方式接到 mRNA5' 端所形成（图 4－23）。在未成熟的 mRNA 的 5' 端都接有一个三磷酸，通常与末端核苷酸（A）相连，在加帽过程中，这个 5' －三磷酸水解，形成 5' －二磷酸，然后在加帽酶（也叫鸟苷酸转移酶，guanylyl trnasferase）作用下，鸟苷酸基团再连接到 mRNA 的 5' 端，最后鸟苷酸基团在一种特殊甲基转移酶的作用下，在第 7 位上甲基化。一般认为，帽子结构具有促进核糖体

体与 mRNA 结合的作用，但是在离体条件下，帽子结构对 mRNA 的翻译并不重要。可能在活体内，帽子结构通过与某种具有特殊功能的蛋白质结合，促进蛋白质合成的启始过程。

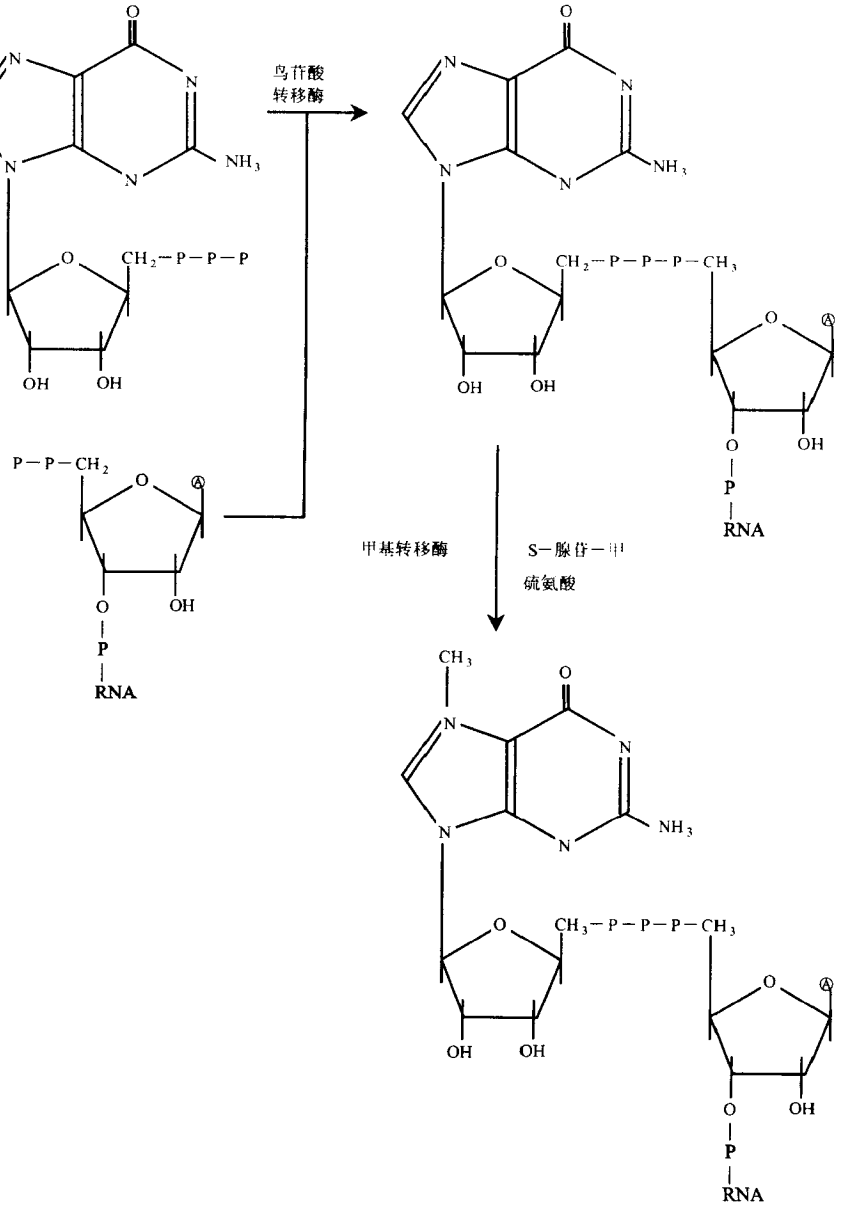


图 4－23　真核生物 mRNA 的加帽过程

（二）mRNA3’端的多聚A尾巴形成

如前所述，几乎所有细胞核基因编码的mRNA的3’端都有一段长达200nt的多聚A尾巴，这种尾巴并不由基因本身编码，而是在转录之后，立即由细胞核中的多聚A聚合酶（polyA polymerase）加到mRNA上的。根据对酵母mRNA加尾过程的研究，多聚A不是直接加到转录终止位点下游，而是加到其上游（图4－24）。加尾位点由一种特殊的内切核酸酶切割而成，这种酶识别加尾信号AAUAAA序列，当结合到mRNA上后，在加尾信号序列下游切掉11～13个碱基，然后再加尾。

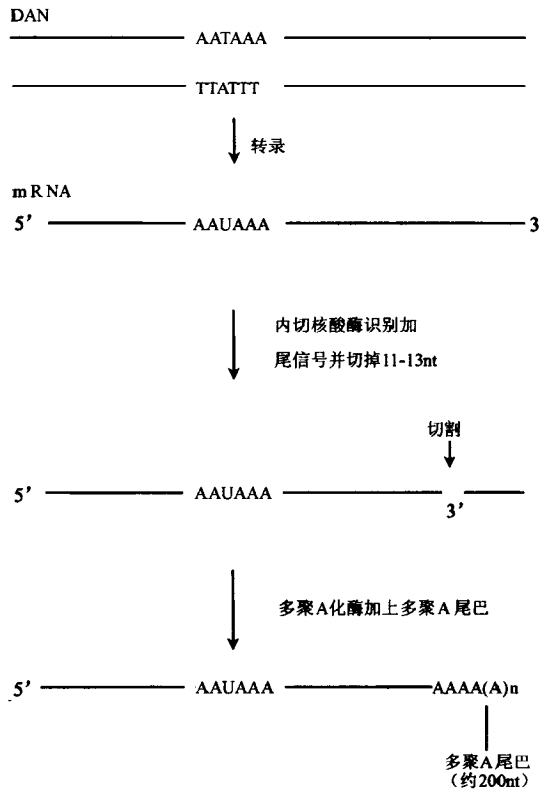


图4－24　真核生物mRNA的加尾过程

内切核酸酶首先在加尾信号的下游切掉11～13个碱基，然后再加上多聚A尾巴

（三）组蛋白mRNA3’末端的加工

真核细胞的组蛋白mRNA同其它大多数mRNA不同，它不具有多聚A尾巴，也不具有AAUAAA加尾信号。但是，在各种真核生物中，组蛋白mRNA

的3'末端附近都有一种高度保守的回纹对称序列（图4-25），这段序列能在接近mRNA3'端的位置形成发夹环结构。这种结构保守序列可能是加工信号而不是转录终止信号。根据对滑爪蟾H_2A基因逐步缺失突变的研究，这种组蛋白mRNA形成正确的3'末端需要DNA分子中相当于mRNA3'末端下游的、长80 bp的序列，其中包括两个相邻组蛋白基因之间一部分间隔序列，其同感序列为CAAGAAAGA。

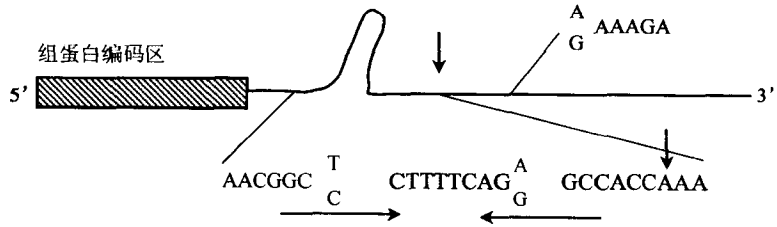


图4-25　组蛋白前体mRNA的3'部分

示两段在海胆和哺乳动物中都具保守性的同感序列，箭头示切割位点

组蛋白mRNA3'末端的正确加工还需要一种小的snRNP，这种复合体含有一种长60nt的、叫做U_7的RNA。根据U_7RNA的序列推测，它可能通过与成熟组蛋白mRNA3'端位点前后的保守序列进行碱基配对而与前体mRNA进行互作。在海胆中，虽然5种组蛋白的前体mRNA都具有几乎相同的发夹结构和CAAGAAAGA侧翼序列，但是其卵母细胞只能加工4种组蛋白前体mRNA（H_2A、H_2B、H_4和H_1），而不能加工H_3，其原因尚有待研究。

（四）内含子的切割

1. *内含子的结构特征*　各种真核生物基因的内含子都有一些共同的结构特征。内含子的两端即与外显子交界处分别具有两个高度保守的碱基，5'端为GT（mRNA为GU），3'端为AG，只有少数基因例外，这类基因的内含子的5'端为AT，而不是GT，3'端为AC而不是AG。内含子通常都由富含碱基AT的序列组成，在内含子两端的GU和AG周围各有一组特殊的结构，即5'－AG/GU$^{A}_{G}$AGU3'和5'（U或C）nN（C或U）AG/G3'（"/"为切割位点），这两种结构保守序列都是mRNA加工的信号序列。此外，在内含子的3'剪接位点附近还有另一段特殊序列UACUA$\underline{A}$C（"－"为RNA的分支点碱基），叫分支点序列（branch point sequence），与内含子在剪接过程中形成套环结构有关（图4-26）。

在某些多内含子基因中，几个内含子按性质不同，将DNA序列分成几个不同功能部分，每一部分都编码同一蛋白质的不同功能结构。但在大多数情况下，内含子的分布与基因的功能部分所在的位置没有直接关系。有些基因的某些内含子还具有可读框，编码某种参与自身的前体mRNA或其他前体mRNA

加工过程的蛋白质。如酵母线粒体细胞色素氧化酶亚基Ⅰ基因的第一个内含子编码一种具有酶性质的蛋白质，这种蛋白质参与自身 mRNA 的加工过程，如果这个内含子中发生突变，酵母细胞就不能合成具有功能的细胞色素氧化酶亚基Ⅰ。有些基因的内含子还位于某个氨基酸密码之内，将一个三联密码隔开，因此内含子的切割必须非常准确。

2．前体 mRNA 剪接机制　前体 mRNA 剪接（pre－mRNA splicing）是将内含子准确地除掉和将外显子连接在一起的过程。剪接过程通过两个转酯基步骤完成（图 4－27）。在第一步切割连接反应中，分支点腺嘌呤基团的 2′－OH 连接到 5′剪接位点的磷酸基上，释放 5′外显子和形成套环结构。套环结构在内含子的 5′端和分支点腺嘌呤之间有一个异常的5′－2′磷酸二酯键。在第二个切割连接步骤中，5′外显子的 3′羟基连接到 3′剪接位点的磷酸基上，使 2 个外显子连接在一起，并释放套环形式的内含子。

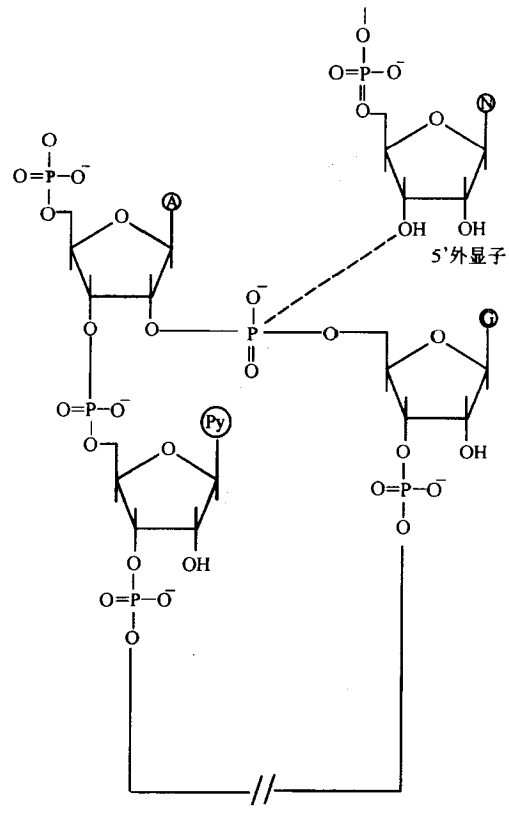


图 4－26　前体 mRNA 内含子中分支点的结构

前体 mRNA 的剪接过程是一种由 5 种 UsnRNA（U_1、U_2、U_4、U_5 和 U_6）催化的反应，发生在一个很大的核蛋白复合体即剪接体（spliceosome）中。5 种 UsnRNA 是这种剪接机器的关键成分，所以它们又被称为核酶（ribozyme）。每一种 snRNA 都由蛋白质包裹成核蛋白颗粒（snRNP），它们在前体 mRNA 上按步骤、有顺序地聚集在一起。在聚集过程中许多步骤都要水解 ATP。首先，U_1 与 5′和 3′剪接位点周围的序列进行碱基互补配对，U_2 与分支点序列结合。然后，$U_4－U_5－U_6$ 三聚体 RNP 结合进来，U_6 取代 U_1 与 5′剪接位点的序列结合，U_2 与 U_6 的部分区域进行碱基互补配对，U_5 形成茎环结构并与 5′和 3′外显子互作（图 4－28）。最后，释放 U_1snRNP，开始第一步切割连接反应。对于少数 5′端和 3′端分别为 AT 和 AC 的内含子而言，其

识别机制可能涉及一些不同的识别因子如 U_{11} 和 U_{12}snRNA 等，但其总的切割连接过程大致相同。

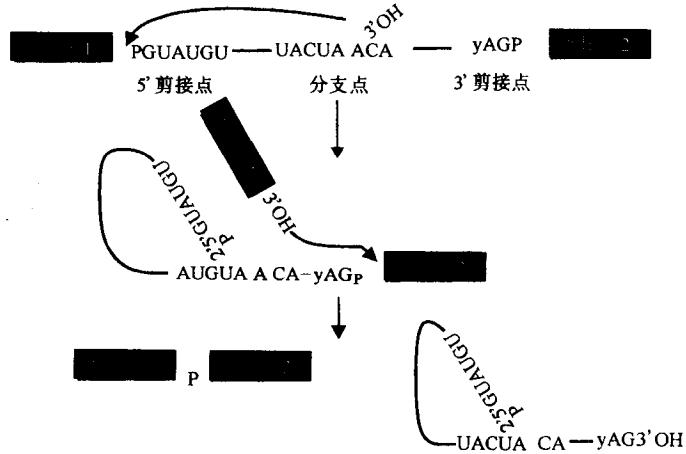
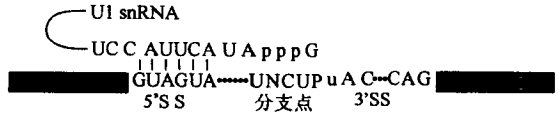


图 4－27　细胞核前体 mRNA 剪接的化学途径

较大的大写字母为内含子中衡定的碱基，较小的大写字母为保守性程度较低的序列

（根据 H.D.Madhani 和 C.Gurthrie，1994）



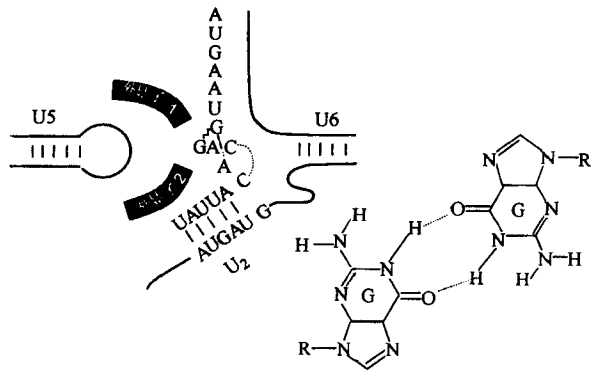


图 4－28　各种 snRNP 与前体 mRNA 互作的模型

－为 5'－2'磷酸二酯键，··为非 Watson－Crick 互作

五、蛋白质合成的启始

（一）翻译启始因子及启始复合体形成

所有生物中的翻译过程都几乎相同，真核和原核生物翻译的最大差别在于启始过程。如前所述，大肠杆菌的翻译启始只需要三种翻译启始因子，但是真核生物则需要大量翻译启始因子 eIF（eucaryotic initiation factor）。这些因子在蛋白质合成过程中的作用如图 4－29 所示。在翻译启始过程中，核糖体小亚基（40s）首先与 mRNA 和启始 tRNA 结合，然后小亚基向 3’端移动，直到找到起始位点，即蛋白质编码序列中第一个 AUG 密码所在位置，这就是所谓的启始复合体形成的扫描模型（scanning model）。最后 60S 大亚基结合进来，形成 80S 复合体。80S 启始复合体是翻译启始反应的最终产物，这一反应过程是不可逆的。80S 核糖体具有一个开放的 tRNA 结合位点（A 位点），因此能进入蛋白质合成的延长期。

真核生物翻译启始过程的一些基本步骤也都与原核生物有所不同，主要表现在：

1．真核生物的翻译启始 tRNA 为 Met－$tRNA_i^{met}$，它不需要甲酰基化。

2．启始 met－$tRNA_i^{met}$和 GTP 与 eIF2 形成一种可以分离的复合体，这一过程不依赖于核糖体小亚基。在原核生物中，IF_2 则与 30S 小亚基上的启始 tRNA 结合。

3．启始 tRNA 在与 mRNA 结合之前就与 40S 核糖体小亚基结合。

4．mRNA 结合进来时需要将 ATP 水解成 ADP。

5．启始过程不仅需要更多的启始因子，而且许多 eIF 因子本身也都是多亚基蛋白。例如，与 mRNA 结合有关的 eIF_3 约有 10 个亚基，分子量为 10^6kd，相当于 40S 亚基的三分之一。

6．启始过程需要 mRNA5′端的帽子结构。

在上述各种差异中，关键的差别可能就是 eIF_2 因子通过循环，进行 GTP－GDP相互转换。循环过程需要另一种叫做 eIF_2B（又叫 GTP－GDP 互换因子，GEF）的因子。eIF_2 循环的各个步骤都由酶催化，与原核生物中肽链延长的 EF－Tu/Ts 循环相似。当 eIF_2 将启始 tRNA 置于核糖体上之后，就以一种非活性形式释放下来，并含有 GDP。当 eIF_2 与 eIF_2B 互作，形成 eIF_2－eIF_2B复合体时，GDP 才被替换下来。在这种复合体中，eIF_2 对 GTP 亲和性提高，因此就能再次进入活化池，并再次与 Met－$tRNA_i^{met}$ 结合。这种磷酸化的 eIF_2 与 eIF2B 互作比较稳定，所以这两种因子就形成 eIF_2－eIF_2B 复合体。

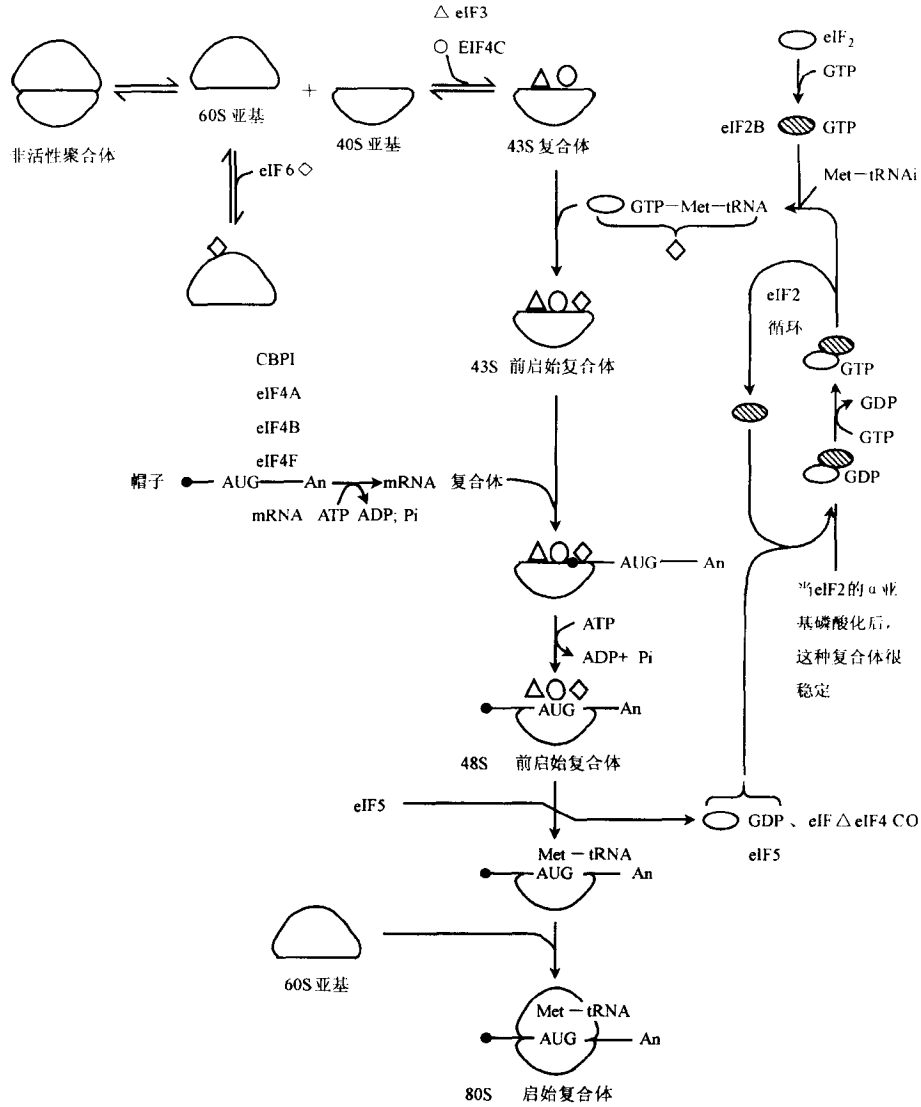


图 4－29　真核生物蛋白质合成启始

eIF2－eIF2B 循环和多种因子与 mRNA 结合，真核生物中 eIF3 和 eIF6 共同执行原核生物 IF3 的一种功能，核糖体分离成大小两个亚基，eIF6 结合到 60S 大亚基上，而 eIF3 则结合到 40S 小亚基上；CPB 为帽子结合蛋白，eIF4F 也叫 CPBⅡ

根据这种模型，核糖体与 mRNA5′ 端结合一次就合成一条多肽链，所以多肽链的产量完全依赖于 40S 小亚基与 mRNA5′ 端的结合速率。当然，结合

速率也受核糖体移动速度的制约，因为前面的核糖体不向前移动，后面的核糖体就不能顺利结合进来。

（二）翻译起点的选择

1. *核糖体与 mRNA5' 末端的结合* 翻译效率的高低首先取决于核糖体与 mRNA5' 端结合的情况，但这一过程目前知道得很少，目前只是对 5' 帽子结构的效应研究得较多。在离体条件下进行翻译时，帽子结构可以促进 40S 复合体的形成。但是在某些情况下，当 mRNA 的 5' 非翻译区发生大片段缺失后，也并不影响翻译的效率。有些实验证明，核糖体与 mRNA5' 端结合需要一种帽子结合蛋白（CBPI），当 CBPI 结合到帽子上后，就促进其他启始因子结合到 mRNA 的前导序列上。其中有一种启始因子 eIF4A 同其他启始因子一起，利用水解 ATP 产生的能量，将起点密码上游的 mRNA 次级结构解螺旋，促使 40S 的小亚基结合进来（图 4－29）。一般认为，40S 小亚基与 mRNA 结合时，它直接穿到 mRNA 上，因为只有具有游离 5' 端的 mRNA 才能进行翻译。真核生物核糖体完全不能启始环状 mRNA 分子的翻译，而原核生物核糖体则能相当正常地翻译环状 mRNA。

2. *起点密码的选择* 真核生物的蛋白质合成启始只有一个方面比原核生物简单，即起点密码的选择。绝大多数真核细胞 mRNA 的起点密码都是 AUG，因为 AUG 距 mRNA5' 端最近。根据扫描模型，40S 小亚基与 mRNA5' 端结合和搜寻第一个 AUG 的过程需要 ATP。当 40S 小亚基到达 AUG 和 Met－$tRNA_i^{met}$ 的互作位点时，便停止滑动，因为 Met－$tRNA_i^{met}$ 在 40S 亚基结合进来之前就已与 mRNA 结合（图 4－29）。Met－$tRNA_i^{met}$ 对 AUG 的准确识别与 $tRNA_i^{met}$ 反密码子 3' 端的一组受修饰的碱基 A 有关，这种修饰作用使密码子－反密码子互作的第一位碱基不发生摆动。在大肠杆菌中，由于 $tRNA_i^{fmet}$ 不具有这种被修饰的碱基，故启始 $tRNA_i^{fmet}$ 与起点密码互作时可逆性很高，它可以识别 GUG、UUG 和 AUU。但是真核细胞 mRNA 并不一定需要类似原核生物 mRNA 上游的额外信号序列，原核生物 mRNA 通过这段序列与 16S rRNA3' 端序列进行碱基配对。

真核生物 mRNA 的起点密码并不是第一个 AUG，而是其下游的一个 AUG。根据对各种 mRNA 的序列分析表明，起点密码的选择可能与 AUG5' 端其他一些核苷酸有关，特别是 －3 位的碱基，该位点几乎全部为嘌呤碱基，其中 80% 为 A。在原核生物中，－3 位的碱基也绝大多数为 A。在植物中，mRNA－3 位和－13 位上也都是 A，而且其下游的 +4 和 +5 位则分别为 G 和 C，这种保守性也可能与核糖体识别翻译起点有关。

根据对酵母细胞色素 C 合成过程的研究，正常情况下，起点密码为 AUG

时，蛋白质合成水平的变化只相差2倍；如果将－3位的A变成U，当起点密码为UUG时，合成水平则下降20倍。在猴细胞中，当mRNA－3位的碱基为A时，其蛋白质合成水平是该位点上为G的3倍，C的15倍；如果将AUG后的碱基A变成G，蛋白质合成水平则提高50%。又如根据对小麦种质细胞在离体条件下核糖体结合过程的研究，当－3位上的碱基为A或G时，其结合效率相同；如果该位碱基为C，核糖体结合效率则降低2～8倍；如果AUG后的碱基为G，核糖体结合水平将比为A时提高1.5倍，比为C时提高2倍，比为U时提高4～10倍；表明核糖体与mRNA结合对AUG周围的碱基有一定倾向性。

另外，真核生物mRNA的起点密码除了AUG以外，其他一些密码也可能用作翻译起点密码。例如在酵母中，mRNA在离体系统中进行翻译时，往往可以在AUU，UUG和AUA上形成80S复合体，而且UUG和AUA可以在活体内启始酵母mRNA的翻译。根据分析，酵母中UUG和AUA只启始低含量蛋白质的合成，由二者启始合成的蛋白质分别占由AUG启始合成的蛋白质的7%和0.4%。

六、翻译延长和终止

当80S的翻译启始复合体形成之后，翻译进入延长期。真核细胞中翻译延长和终止所需的因子相当简单（表4－3）。其中EF1α相当于原核生物的EF－Tu，其分子量为50kd，与GTP和氨酰基－tRNA形成复合体，并将后者传递到核糖体上。$EF_1\beta\gamma$催化GDP－GTP的相互转换，并为氨酰基－tRNA结合过程的下一轮循环准备$EF_1\alpha$因子。真核生物EF_2的分子量约为100kd，相当原核生物的EF－G因子，水解GTP并催化氨酰基－tRNA随mRNA的移动从核糖体A位移到P位。

表4－3　真核细胞中与肽链延长和终止有关的因子

因　子	功　能	原核生物相当因子
$EF_1\alpha$	将氨酰基tRNA结合到核糖体上	EF－Tu
$EF_1\beta\gamma$	协助EF1α再循环	EF－Td
EF_2	移位	EF－G
RF	肽链释放	RF2

肽链合成的终止过程只涉及一种叫做RF的因子，分子量为115kd。这种终止因子具有原核生物RF1和RF2所具有的作用，能识别所有三种终止密码UAA、UGA和UAG。当RF因子活化了翻译肽链的肽基转移酶（peptidyl

transferase）后，翻译终止过程需要水解 GTP，使 RF 因子脱离核糖体。另外，高等真核细胞中，各种氨酰基－tRNA合成酶通常以聚合体形式存在，具有多种合成酶活性，并含有其他一些成分，它可能是一种与膜结合的酶类。在哺乳动物中，其主要功能之一可能是使细胞直接将外源氨基酸整合到蛋白质之中，使之不与细胞内的大量氨基酸混合。

第四节　细胞核基因表达的控制

在控制基因活性的机制方面，真核生物和细菌非常相似。无论是原核生物还是真核生物，反作用子首先与顺作用子结合，使基因的转录活性改变，这些正、负转录因子分别提高或抑制 RNA 聚合酶对启动子的识别能力。由于转录是基因表达的第一步，可能真核生物中大多数基因的表达都是在转录水平上进行调节。但是基因表达在其他水平上的控制如 RNA 加工、翻译、蛋白质加工、运输和降解等也都会对细胞的功能产生重要影响。

一、基因表达在转录和转录后水平上的控制

（一）多启动子与可变转录

同细菌一样，真核生物的许多基因都有 2 个启动子。在这些启动子中，有一个为组成型，其他则为诱导型启动子，因此生物体在控制基因表达方面就可以获得额外的可塑性。例如，在小鼠和家鼠的腮腺（parotid gland）中，α－淀粉酶基因（Amy－1）从上游启动子启始的 mRNA 的合成量是从下游启动子启始的 mRNA 的合成量的 100 倍（图 4－30A）。产生这种差别的主要原因是两个启动子的强度不同。在这两种动物的腮腺腺泡细胞中，上游启动子 P_1 的活性特别高，而 P_2 只是在肝细胞中才具有活性，在小鼠的腮腺和胰脏中，P_1 和 P_2 则同时活动。

除了细胞特异性以外，小鼠的 Amy1 基因的两个启动子在腮腺分化期间还具有不同阶段活性特点。小鼠在出生时就可测定到 P_2 转录子，出生后两个星期，其浓度增加，直到达到成熟个体的水平。另一方面，在小鼠出生后的第一周内测定不到 P_1 转录子，但在出生后第二至第三周内，P_1 转录子上升 100 倍。这一时期正是小鼠和家鼠腮腺的腺泡细胞进行增殖和分化的时期。因此，在未分化的腺泡细胞中，可能 Amy1 的 P_2 启动子由一般性转录因子活化，而在此后，可能 P_1 由腮腺特异性的转录因子活化。

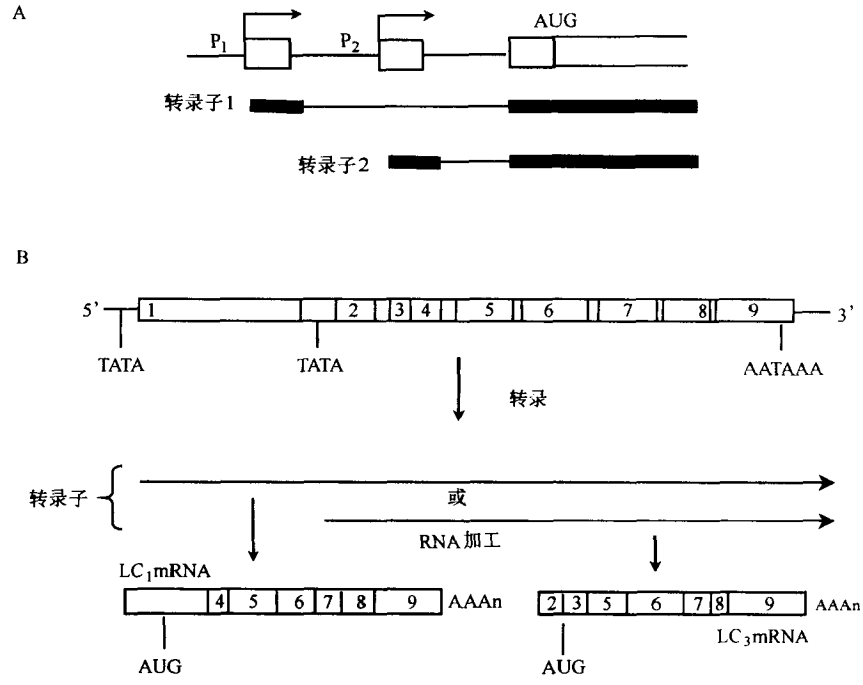


图 4－30　通过多启动子可变转录和转录子可变剪接产生编码不同蛋白质的 mRNA

A. 小鼠 α－淀粉酶基因（Amy－1）多启动子的可变转录，粗线示成熟 mRNA 的序列　B. 鸡的肌球蛋白基因通过可变转录和转录子可变剪接产生不同的 mRNA，阿拉伯数字为外显子

通常蛋白质的初级结构并不受转录起点的影响，因为翻译启点密码通常位于外显子区内。但也有一些例外情况，使用多启动子可使基因表达具有质的差别。

高等真核生物能够利用同一基因在不同组织中编码不同蛋白质，如果某个基因上游存在两个启动子，每一启动子又具有自身的组织特异性调节序列，那么该基因就能产生两种依赖于组织特异性的、5’端长度不等的转录子。如果转录子的 5’区域中含有剪接信号和编码区，那么在有关组织中就合成两种 N－末端氨基酸序列不同的蛋白质。例如在鸡基因组中，肌球蛋白（myosin）基因采取这种表达方式产生两种不同的肌球蛋白轻链。心肌含有一种轻链叫做 LC_1，沙囊含有另一种轻链叫做 LC_3，而骨骼肌则含有 LC_1 和 LC_3 两种轻链（图 4－29B）。该基因的两个相距 10kb 的启动子决定两种前体 mRNA 的合成，可能这两种初级转录子再经过可变剪接，形成 LC_1 和 LC_3 两种 mRNA。

高等真核生物中，采用多启动子控制可变转录的基因还有很多，特别是一

些受发育调节的基因更是如此，如果蝇的酒精脱氢酶基因（Adh）、家鼠和人体的醛缩酶基因（aldolase A gene）以及果蝇的同源异型（homeotic gene）基因等。

（二）转录子的可变剪接产生不同的 mRNA

根据对许多基因表达过程的研究知道，一个基因的同一种转录子在不同组织中通过不同加工途径可以产生具有不同编码功能的 mRNA。如家鼠降钙素基因的转录子就属于这种情况。在甲状腺细胞中，其转录子在上游的一个位点上多聚 A 化，而在大脑细胞中，则在下游的一个位点上多聚 A 化（图 4－31）。同前述肌球蛋白转录子的加工一样，降钙素基因的两种 mRNA 也是通过可变剪接途径形成的。降钙素基因表达有可能是通过选择剪接位点进行控制，也有可能通过选择多聚 A 化位点进行控制。如果多聚 A 化发生在剪接之前，细胞

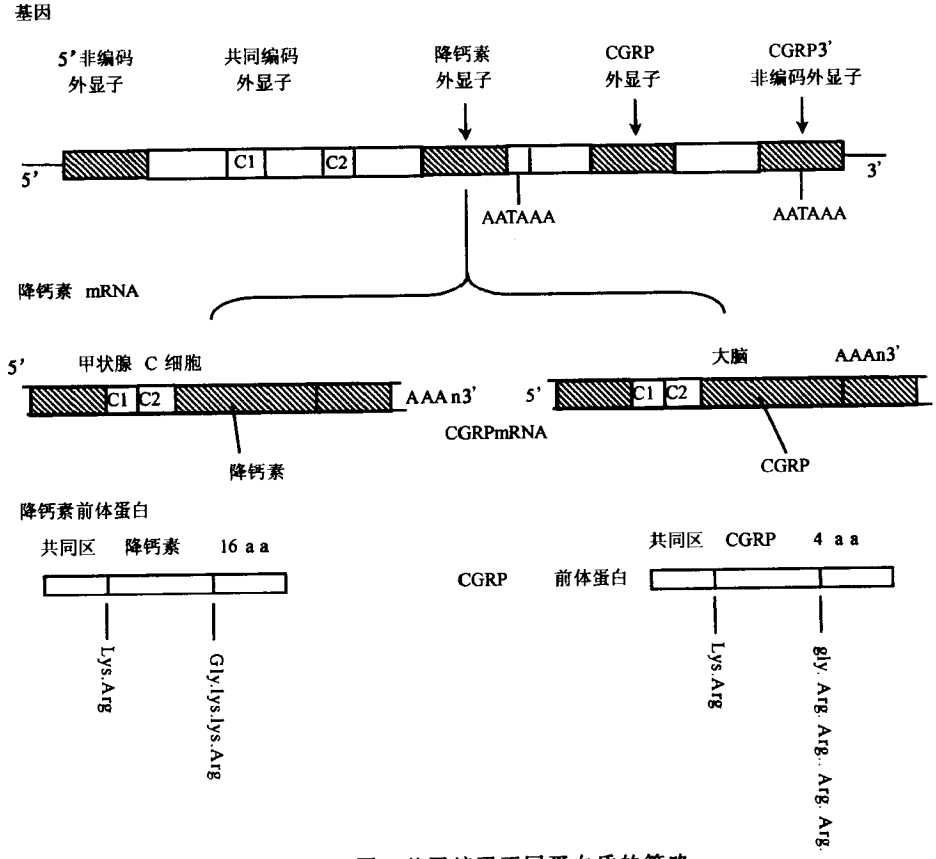


图 4－31　同一基因编码不同蛋白质的策略

示降血钙素基因的转录子通过不同加工途径形成降血钙素和一种神经肽，叫做降钙素有关的肽（CGRP）

就以组织特异性方式控制多聚 A 化，前体 mRNA 就含有可供选择的剪接位点。如果初级转录子在合成后就立即进行剪接，剪接后的转录子的结构就会影响多聚 A 化位点的选择。但是，无论在任何情况下，最后都会形成两种具有三个相同外显子和不同 3′外显子的 mRNA，从而在两种细胞中翻译成不同多肽。这两种前体多肽最后又通过加工，形成不同激素。

基因的同一转录子通过可变剪接途径可以形成两种编码功能完全相反的 mRNA，特别是一些编码调节因子的基因，这些基因通过对自身转录子的可变剪接，指导合成调节其他基因表达的活化子或抑制子。例如，原癌基因（proto－oncogene）erbAa 编码一种细胞核甲状腺激素受体的亚型蛋白，甲状腺激素与这种受体结合，然后再与启动子区中不同成员结合，抑制或活化被调节基因的转录过程。

由 erbAa 基因编码的受体蛋白的氨基末端部分含有 DNA 结合区，其羧基末端则含有配位体（甲状腺激素）的结合区。这一基因通过可变剪接形成 α_1、α_2 和 α_3 三种不同的转录子。从这三种转录子翻译成的蛋白质在长 370 个氨基酸的 N 基末端部分完全相同，只是 C 基末端部分有差别。根据对这三种蛋白质功能的分析，α_2 和 α_3 蛋白都不能与甲状腺激素结合，但都能与靶 DNA 序列结合，α_1 转录子单独表达能活化由该激素诱导的转录活化作用，α_1 和 α_2 转录子联合表达则抑制由该激素诱导的转录活化作用。显然，可变剪接机制可以通过影响蛋白质的羧基末端氨基酸序列来修饰受体的配位体特性及其活性程度，从而决定诱导因素活化或抑制有关基因转录。

（三）细胞周期对组蛋白基因表达的调节

在细胞周期的 S 期，真核细胞需要大量组蛋白分子将新合成的 DNA 分子包裹成核小体，而细胞几乎只在 S 期合成组蛋白。组蛋白合成同 DNA 复制是直接偶联在一起的，如果抑制 DNA 合成，则同时终止了组蛋白合成。组蛋白基因表达是在转录水平和转录后水平上受细胞周期调节的。

有丝分裂后，细胞中几乎不存在任何组蛋白 mRNA，当细胞进入 S 期时，细胞中组蛋白 mRNA 含量急剧增加。组蛋白基因的转录与 S 期特异性转录因子有关，这些因子有选择地活化组蛋白基因的启动子。此外，组蛋白 mRNA 在合成之后，依据细胞周期的阶段，命运也会不同。由于组蛋白 mRNA 不具有多聚 A 尾巴，半衰期（half life）极短，只在 S 期其半衰期才延长，当 S 期完成后，组蛋白 mRNA 迅速降解。

（四）mRNA 的寿命与转录控制

在真核细胞中，许多 mRNA 的寿命非常有限，在生长的脊椎动物细胞中，mRNA 的半衰期平均为 3 小时。虽然细胞中许多 mRNA 在整个生命期间都能

始终如一地进行翻译，即合成所谓的“看家蛋白质”，但是还有许多重要蛋白质或 RNA 合成过程的启动或关闭必须依照细胞周期的规律性和严格顺序进行。例如在酵母中，大约有 50 个基因控制细胞分裂周期（CDC），如果这些 cdc 基因中任何一个受到高温破坏，都会使细胞发育停止在某个特定阶段。可能其中至少有某些基因编码不稳定 mRNA，这些 mRNA 的短暂存在都能使细胞从细胞周期的一个阶段进入另一阶段。

相反，在高等动物的最终分化的细胞中，许多 mRNA 是相当稳定的。例如网织红细胞在合成其主要蛋白质血红蛋白时，实际上并不合成 RNA。网织红细胞合成的血红蛋白占其整个蛋白质含量的 90%，因此就无须破坏血红蛋白 mRNA 而再去合成新的 mRNA 分子。同样，在成熟个体的肝细胞中，大多数 mRNA 都编码血浆蛋白（plasma protein），这些蛋白质以后又都被释放到循环系统中去，因此肝细胞中的 mRNA 也是相当稳定的。

mRNA 的半衰期与许多因素有关。虽然对 mRNA 降解的详细过程还不清楚，但现在知道，mRNA 稳定性与自身分子内许多序列有关，包括 5′帽子结构、5′非翻译区、可读框中某些特殊序列、无义突变的终止密码、3′非翻译序列以及多聚 A 尾巴等。由于这些序列能够解除 mRNA 的稳定性，调节 mRNA 的降解过程，所以它们统称为 mRNA 的脱稳定子（destablizing element）。在这些特殊序列中，mRNA 的多聚 A 尾巴是最重要的调节子之一。根据研究，mRNA 只有在其多聚 A 尾巴缩短到 10 个碱基以下时才会降解。一般情况下，mRNA 在翻译后，其多聚 A 尾巴都与一些特殊的调节多聚 A 尾巴缩短过程的蛋白质结合。

现在已从酵母中分离到一种缩短多聚 A 尾巴的特殊核酸酶叫做多聚 A 核酸酶（poly A ribonuclease，PAN），PAN 酶并不能作用于裸露的 RNA，它是一种依赖于多聚 A 结合蛋白的特殊 RNA 酶。这种酶缩短多聚 A 尾巴的作用或脱末端腺嘌呤的作用受 mRNA3′非翻译序列的调节。此外，PAN 酶还是一种翻译启始因子。

有关 mRNA 中脱稳定序列的作用机制目前还不清楚，可能这些序列只有通过翻译过程才能行使功能，因为只有通过翻译机器才能将某些调节因子带到 mRNA 上。各种脱稳定序列被 mRNA 降解调节因子识别以后，它们可能通过 3′末端脱腺嘌呤步骤诱导 mRNA 退化。末端脱腺嘌呤步骤完成之后，外切核酸酶和内切核酸酶识别 mRNA 分子内脱稳定序列或其他有关序列，破坏 mRNA（图 4－32）。

另一方面，mRNA 的半衰期可能与某些特别稳定的蛋白质与 mRNA 结合有关，这类蛋白质可能在基因表达过程中行使一种正调节因子的功能。例如，

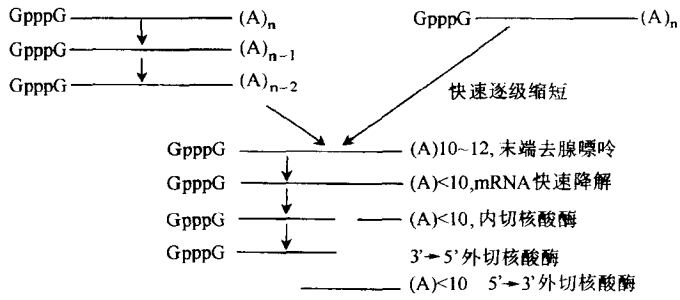


图 4－32　去嘌呤作用与 mRNA 降解

正常的破坏多聚 A 的核酸酶逐级缩短多聚 A 尾巴，这一过程由 mRNA 的脱稳定序列所调节。一旦 mRNA 去腺嘌呤之后，各种核酸酶都可以破坏 mRNA

虽然某些无义突变可以解除 mRNA 的稳定性，加速 mRNA 降解，但是根据对酵母中抑制组氨酸代谢基因 his4 的 mRNA 无义突变的抑制子的研究，其中有一种叫做 UPF_1 的抑制蛋白能够稳定含有无义密码的 mRNA。可能在不同组织或不同细胞中存在某些特殊的、能稳定 mRNA 的正调节因子。

（五）5SrRNA 基因表达的自动调节

根据在离体条件下对 RNA 聚合酶Ⅲ转录过程的研究，各种转录因子与 DNA 之间形成稳定转录复合体后，能多轮次地进行 RNA 合成。如前所述，滑爪蟾基因组含有 2 套不同的编码 5SrRNA 的基因，其中有一组具有卵母细胞表达特点，只在卵子形成期间表达。因此，卵细胞贮存了各种大分子化合物，以用于受精后合子的发育。体细胞特异性表达的 5SrRNA 基因通常只在受精后开始表达，并一直延续到滑爪蟾的整个生命过程。

卵母细胞特异性 5SrRNA 基因转录除了需要 RNA 聚合酶Ⅲ以外，还需要另外 3 种因子。其中有两种即 TFⅢB 和 TFⅢC 为普通转录因子，二者除了能调节 5SrRNA 基因转录外，还能调节其他基因如 tRNA 基因的转录。而第三种转录因子 TFⅢA 则为一种卵母细胞特异性蛋白（40kd），它与基因的内部转录控制信号紧密结合。当未成熟的卵母细胞合成了足量的 TFⅢA 之后，这种转录因子即与卵母细胞特异性的 5SrRNA 基因结合，活化 RNA 聚合酶Ⅲ，启始该基因转录（图 5－33）。

过去曾认为卵母细胞中贮存的特异性 5SrRNA 是一种与特殊蛋白质结合的复合体，以 RNP 的形式存在，后来发现这种蛋白质与 TFⅢA 属同一种蛋白质，因此，当卵母细胞合成足量的 5SrRNA 之后，就由一种极其精确的机制关闭卵母细胞特异基因的表达。在卵细胞形成的初期，当 TFⅢA 多于 5SrRNA 时，这种调节蛋白就与 5SrRNA 基因结合，促进其转录。在卵子形成的后期，随着

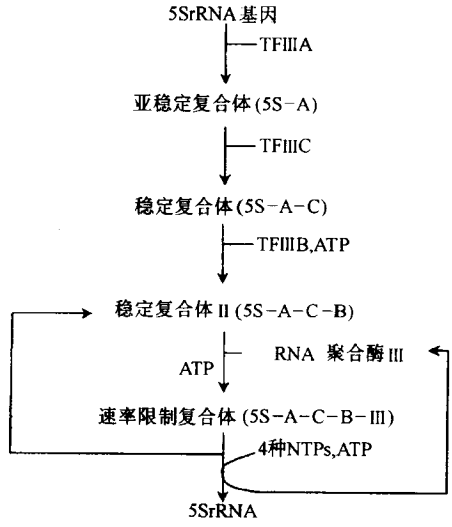


图 4-33　5SrRNA 基因与转录因子互作示意图

5SrRNA 的累积，TRⅢA 以 5SRNP 的形式被分开，使 5SrRNA 合成成比例地下降。最后所有 TFⅢA 均以贮存颗粒的形式贮存在细胞之中，5SrRNA 完全停止合成。受精之后，5SrRNA 用于构建核糖体，TFⅢA 开始降解。显然，5SrRNA 基因的转录是 TFⅢA 与 DNA 结合的直接结果。

二、基因表达在翻译水平上的调节

mRNA 在转录后，在细胞核内通过加工，然后再运输到细胞质中，将基因组所含遗传信息翻译成蛋白质。基因在翻译水平上的控制表达是生物体发育的一种重要调控机制，mRNA 是否进行翻译取决于细胞内的特殊环境条件。因此，当细胞急需合成某种蛋白质时，基因就可在翻译水平上控制表达。另一方面，在翻译水平上控制基因表达还是一种控制蛋白质合成的微调机制，它根据细胞内已存在的 mRNA，合成生长发育所需的蛋白质。基因在翻译水平上的调节表达可能是控制发育的一种主要机制。以下讨论的各种蛋白质合成过程只是部分已知例子，可能还有许多其他基因的翻译调节机制目前还不清楚。

（一）蛋白质协调合成的翻译控制

遗传调节的主要问题之一是在基因组的不同区域协调合成几种产物。在各种不同真核细胞中，蛋白质合成的协调控制也与原核生物相似。例如血液中，血红蛋白由 2 条 α- 和 2 条 β- 珠蛋白链以及 4 个血红素分子组成，当一个红血

细胞合成血红蛋白时，它必须使 α－珠蛋白链、β－珠蛋白链和血红素分子严格地按 2:2:4 的比例合成，任何偏离这种比例的情况都会导致严重疾病。

现在知道，血红素分子调节血红蛋白各种成分按比例地合成。调节方式有两种：第一，过量血红素通过抑制 σ－氨基乙酰丙酸合成酶（σ－aminolevulinate synthase DALA）的活性而关闭自身的合成。DALA 合成酶是血红素合成支路中第一种酶（图 4－34），因此，当血红素含量多于与其结合的各种分子的含量时，细胞就停止合成血红素。第二种方式是过量血红素刺激珠蛋白合成。例如，将血红素加到离体翻译系统中时，珠蛋白合成显著提高。如果没有珠蛋白与血红素结合，血红素将关闭自身合成，而刺激合成更多的珠蛋白（图 4－35）。

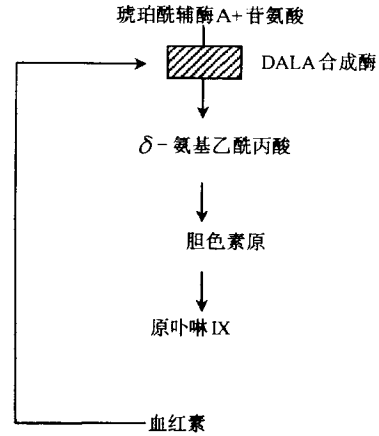


图 4－34　血红素合成的调节循环

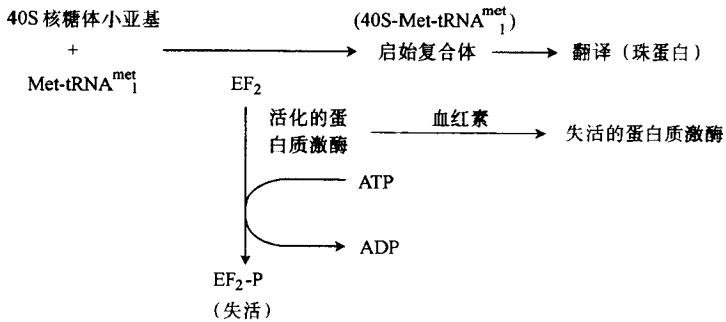


图 4－35　珠蛋白合成在翻译水平上的控制

启始因子（EF$_2$）被蛋白质激酶抑制，只有血红素抑制蛋白质激酶后，EF$_2$ 才具有活性

根据其它许多对血红素调节珠蛋白合成过程的研究，在存在翻译启始因子（EF_2）的条件下，当不存在血红素时，某种特殊的蛋白质激酶使启始因子磷酸化。磷酸化的启始因子并不具有活性，因此就不能使 Met－$tRNA_i^{met}$与 40S 核糖体亚基结合，从而阻止翻译。另一方面，当血红素过量时，过量的血红素能够与蛋白质激酶结合，抑制其活性。失活的激酶就不能使 EF_2 磷酸化，因此就能正常地进行翻译，也就是说，只要存在血红素，就能继续合成珠蛋白。

在二倍体鸡基因组中有 4 个 α－珠蛋白基因和 2 个 β－珠蛋白基因，如果每一基因都以相同速率转录和翻译，那么 α－珠蛋白的合成量就可能是 β－珠蛋白合成量的 2 倍。但事实并非如此，细胞中 α:βmRNA 的比例为 1.4:1，而两种珠蛋白的比例则为 1:1。因此，使这两种蛋白质平衡合成的机制必然发生在翻译水平上。

一般认为，这两种珠蛋白的平衡合成受翻译启始步骤制约，理由是 α－珠蛋白 mRNA 与 β－珠蛋白 mRNA 竞争翻译启始因子，而 β－珠蛋白 mRNA 则又是一个较强的竞争者。启始因子对后者的识别效率较高，所以翻译频率也较高。这两种珠蛋白 mRNA 的 5’序列差别很大，当两者 mRNA 等量存在，且只有有限的翻译启始因子时，翻译产物中 α－ 珠蛋白只占 3%左右。如果将裂解细胞中的总体 mRNA 加到过量启始因子的反应中，这两种蛋白质的比例则为 1.4:1。因此，血红蛋白中 α－和 β－珠蛋白的比例受到翻译启始过程的控制。虽然血红蛋白合成也在转录水平和转录后的加工过程中受到调节，但最后血红蛋白分子的建成则在翻译水平上受到微调。

（二）卵母细胞 mRNA 的翻译控制

在大多数动物中，二倍体合子核中的基因并不马上表达，因为胚胎的早期发育受贮存在卵细胞中或由卵细胞合成的某些物质的控制。在胚胎发生的初期，胚胎主要表现为母本特征，只是在胚胎发育的后期才转向父本或双亲的遗传特征。这种母体影响在锥实螺壳的旋转方向方面早已广为人知，其卵母细胞质含有一种控制卵裂面向左旋还是向右旋的因子。而在海胆中，其杂交种的间叶细胞（mesenchyme cell）数则完全取决于提供卵细胞的亲本种，而父本基因只是在骨细胞形成期间才起作用。

卵细胞中控制胚胎早期发育的贮存 mRNA 只是在受精后才进行翻译，有关这些 mRNA 进行翻译的调节机制目前有多种假说，现以其中三种分别加以说明。

第一种假说认为，卵母细胞 mRNA 由蛋白质包裹，形成 RNP，使其不能与核糖体接触。与 mRNA 结合的蛋白质种类各不相同。由于包裹的 mRNA 不能与核糖体结合，所以就不能翻译。受精后，包裹 mRNA 的蛋白质被除掉，所以 mRNA 就能进行翻译。因此，在胚胎发生的初期，受精卵通过启始翻译过程来

控制基因表达。

第二种假说认为，某些蛾类采用另一种不同机制控制卵细胞的翻译过程。根据研究，烟草天蛾卵母细胞贮存的 mRNA 都有一个未甲基化的帽子结构，即只有鸟嘌呤而无甲基。在离体系统中，这种 mRNA 不能翻译出蛋白质。受精以后，卵细胞中所有 mRNA 都甲基化，形成完整的帽子结构。具有完整帽子结构的 mRNA 就能与核糖体结合和启始翻译过程。

第三种观点认为，卵母细胞中的 RNP 是能够翻译的，只是其蛋白质合成机器被隔开，致使 RNP 中的 mRNA 无机会与核糖体接触。海胆卵母细胞中的组蛋白 mRNA 可能就受这种机制所调节。海胆卵母细胞中的组蛋白 mRNA 不存在于细胞质 RNP 之中，而是位于原核（pronucleus）之中。只有当原核解体后、受精过程结束时，组蛋白 mRNA 才进入细胞质。

（三）tRNA 和 tcRNA 与翻译控制

在细菌和酵母中，各种 tRNA 的相对含量与高度表达的基因中密码的使用情况之间存在很高相关性。在这些生物的基因组中，tRNA 和 mRNA 通过协同进化和优化组合，以最大的速率合成大量蛋白。但是在多细胞生物中，不同组织则合成不同的大量蛋白。在某个特化细胞中，各种 tRNA 的总体比例只满足自身需要，而不适宜另一特殊细胞类型。因此高等真核细胞中存在交替调节 tRNA 群体的机制，使每一种特化组织都只合成具有特殊氨基酸组成的蛋白质。

在家蚕（*Bombyx mori*）的丝腺细胞中，这种特化细胞合成两组多肽，叫做丝纤朊（fibron）和丝胶（sericin）。这两组丝多肽分别富含某些氨基酸，如丝纤朊富含 Gly、Ala 和 Ser，丝胶则富含 Ser、Gly 和 Asp。在丝胶和丝纤朊 mRNA 合成期间，$tRNA^{Gly}$和 $tRNA^{Ala}$的合成提高 30 倍以上，其中包括一种新的、具丝腺特异性的 $tRNA^{Ala}$。虽然目前还不清楚这些 tRNA 基因转录的机制，但可以肯定，其中必然涉及某种组织特异性的转录因子。

在某些特化的组织中，除了通过调节某些 tRNA 基因的转录来控制某些 mRNA 的翻译以外，其基因表达机制还可使用不同的同功 tRNA 来调节翻译过程。各种同功 tRNA 具有不同反密码子，但都只与一种氨基酸结合。如果某个细胞中 90% 的 $tRNA^{ser}$识别 UCU 密码，而只有 1% 的 $tRNA^{ser}$识别 UCC 密码，那么具有 UCU 的 mRNA 就比具有 UCC 密码的 mRNA 的翻译效率高得多。

高等真核细胞中还有另一类 RNA，叫做翻译控制 RNA（translational control RNA，tcRNA），它与翻译过程有关。tcRNA 是一类与 mRNA 结合的 RNA 分子，以两种形式存在。第一类为细胞特异性 tcRNA，它们抑制具有其它细胞特征的 mRNA 的翻译。例如，幼期红血细胞中的 tcRNA 抑制肌球蛋白 mRNA 的翻译，而成肌细胞中的 tcRNA 则抑制珠蛋白 mRNA 的翻译。第二类 tcRNA

存在于 RNP 颗粒之中，它们与蛋白质一起将 mRNA 包裹起来，抑制 mRNA 翻译。这类 tcRNA 可能使成肌细胞将肌球蛋白 mRNA 贮存在 RNP 颗粒之中，直到细胞融合完成之后才将其释放出来。体细胞中的这种翻译控制机制与卵母细胞中的翻译控制机制非常相似。

三、基因表达的翻译后调节

（一）蛋白质在翻译后的变化是基因表达的调节机制之一

蛋白质在合成之后，它可能是细胞骨架的组成成分，也可能参与细胞内某种代谢产物的合成或分解过程。无论是哪一种情况，某种蛋白质都是细胞内复合成分之一，同其他蛋白质都有着不可分割的联系。蛋白质在合成之后要发生几种变化。首先，新合成的蛋白质如果不经过进一步修饰，就不具有活性。第二，某些蛋白质可能有选择地受到抑制。在某些情况下，这种失活过程可能与蛋白质自身的降解过程有关。而在另一些情况下，可能蛋白质通过与某些抑制配位体结合而失活。第三，某些蛋白质必须位于细胞内某些特定位置才具有活性。因为细胞不仅仅只是贮存酶的场所，而且蛋白质还必须进入细胞内某个特定区域如细胞膜、溶酶体（lysosome）、线粒体（mitochondron）、叶绿体（chloroplast）等细胞器之中，在特定的位置行使某种功能。第四，有些蛋白质需要同其他一些蛋白质结合在一起，组装成功能单位，如血红蛋白、微管蛋白（microtubule）和核糖体等都是由多条蛋白质分子结合在一起形成的功能单位。因此，遗传信息的表达仍然能够在翻译后水平上受到影响。

（二）蛋白质在翻译后通过修饰而被活化

许多新合成的多肽都不具有活性，翻译后必须将某些氨基酸基团切除掉才成为活性分子。特别是某些肽激素的活化过程更是如此。例如，胰岛素的前体是一种含有二硫键的很长的多肽，只有将其中间及一部分氨基末端序列切除掉后，剩下的两段末端序列再通过二硫键结合在一起，才形成具有活性的分子（图 4－36）。

有些蛋白质只有与若干条小分子肽链共价结合、受到修饰后才有活性。例如，磷酸化激酶如果不经过磷酸化就无活性，一旦磷酸化后，就具有催化活性。又如胶原蛋白（collagen），这种蛋白质是体细胞的一种主要结构蛋白，只有在其脯氨酸和某些赖氨酸水解以后才能行使功能，如果这两种氨基酸不被水解，则导致严重疾病。

蛋白质在翻译后的修饰作用能够调节自身基因的转录。在多头绒泡菌中，组蛋白 H_4 依其乙酰化状态，以几种不同形式存在。高度乙酰化的 H_4（2－4 酰

基）具有调节自身基因转录的作用。相反，组蛋白 H_4 的磷酸化作用则使染色质浓缩，遏制转录过程。多头绒泡菌 H_4 的乙酰化作用和 H_1 的磷酸化作用具有可逆性，因此染色质的转录在总体水平上可能受组蛋白的磷酸化和乙酰化作用的调节。

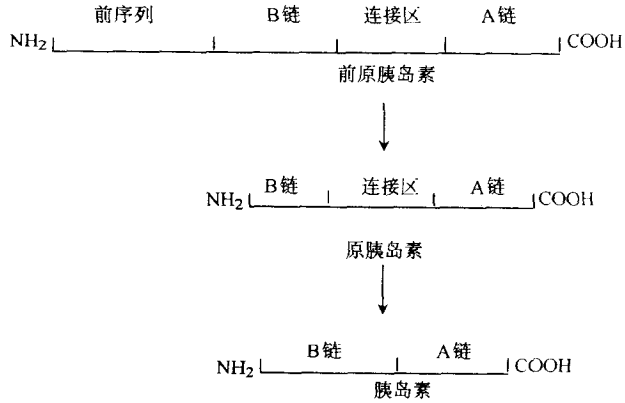


图 4－36　小鼠胰岛素Ⅱ的成熟过程

（三）蛋白质通过翻译后的修饰而失活

在前述血红蛋白的合成过程中，磷酸化作用可以使翻译启始因子 2 失活并以此来调节血红蛋白的合成。同样，磷酸化作用也可以使糖原合成酶（glycogen synthetase）和丙酮酸激酶（pyruvate kinase）失活。而谷氨酰胺合成酶则在接受了一个分子的 AMP 以后而失活。

（四）目标蛋白质的降解

基因表达在转录后调节的另一种方式是在翻译后降解某些特殊多肽。如催化丙酮酸和乳酸可逆互变的乳酸脱氢酶（lactate dehydrogenase LDH），就属于这一范畴。在家鼠中，LDH 由两个不同基因编码，其中一个编码 LDH－A 亚基，另一个编码 LDH－B 亚基，二者位于不同染色体上。具有功能的 LDH 为一种四聚体，由 4 个亚基组成，既可以是亚基 A，也可以是亚基 B。如果这两个亚基基因都表达，就可能形成 5 种不同的 LDH，即 A_4、A_3B_1、A_2B_2、A_1B_3 和 B_4。这 5 种变异体统称为同功酶（isoenzyme），在不同的细胞背景中，它们各具有不同的催化效率。如果这两个基因同等表达，其产物随机组合，这 5 种同功酶的比例为 1（A_4）∶4（A_3B_1）∶6（A_2B_2）∶4（A_1B_3）∶1（B_4）。由于这 5 种同功酶的电荷不同，因此各种酶就可以通过电泳和染色加以区别。但是，在不同器官中，这 5 种同功酶则严重偏离上述理论比例。如肌肉 LDH 同功酶主要由亚基 A 组成，而心脏和肾脏中的同功酶则主要由亚基 B 组成。进一步研

究知道，LDH－A 在各种组织中的合成量差别并不大，但其降解速率则差别很大，在心肌中，其降解速率是骨骼肌中的 20 倍。所以各个 LDH 同功酶基因的表达是在翻译后、酶的降解水平上进行控制，而不是在转录和翻译水平上进行控制。

各种真核生物往往需要利用一种叫做遍在蛋白或泛素（ubiquitin）的蛋白质来确定需要降解的蛋白质，降解过程需要 ATP。由于肽链切割并不是一种耗能反应，可能能量主要消耗在使降解过程特异化的渠道方面。遍在蛋白为一种长 76 个氨基酸的多肽，从酵母到人体，其序列几乎无变化。同由氨酰基－tRNA 合成酶催化的反应一样，遍在蛋白通过一系列活化过程才结合到目标蛋白上（图 4－37）。遍在蛋白在活化后，通常有几条分子结合到蛋白质基质上，在赖氨酸基团的∈－氨基和遍在蛋白羧基末端甘氨酸之间形成异肽键。最后遍在蛋白化的蛋白质通过降解，释放出完整的遍在蛋白分子，以用于下轮蛋白质的降解过程。根据分析，细胞中约有 10％～20％的 H_2A 组蛋白发生遍在蛋白化，这一修饰过程倾向性地发生在转录区的核小体之中，因此，蛋白质有选择地降解是染色质从非活性状态过渡到活性状态的基本步骤。近年来还发现，哺乳动物中依赖于遍在蛋白的蛋白质水解过程需要 tRNA，从而将蛋白质的降解过程同 RNA 联系在一起。

（五）蛋白质的运输

所有细胞不仅必须以一种受控方式合成各种蛋白质，而且还必须将这些蛋白质投放到细胞内、外适当位置。对于原核生物来说，这种蛋白质的投放过程只需通过简单的扩散作用就可达到目的，但对于真核细胞而言，由于细胞很大，被膜丰富，而且每一种被膜都由特殊的蛋白质组成，所以真核细胞必须采取一些特殊运输机制，将蛋白质运输到目标位置。

真核细胞中，许多在内质网膜上由聚核糖体合成的蛋白质大多都须穿越内质网，包括分泌蛋白、溶酶体蛋白、某些膜蛋白以及植物的种子贮藏蛋白等。这些蛋白质在其合成过程尚未结束时，就已开始穿越内质网的移位过程，这种现象称为共翻译分泌（cotranslational secration）。所有进入内质网的蛋白质都有一个共同特征，即都具有一段叫做信号肽的 N－末端序列。信号肽一般长约 15～30 个氨基酸，且都富含疏水氨基酸。它们可能都通过折叠成螺旋状的结构，以便其插入到双层膜之中，启始运输过程（图 4－38）。当增长的肽链最后通过被膜时，信号肽被一种特殊的水解酶叫信号肽酶（signal peptidase）剪掉。蛋白质在进入内质网小泡后，通过加进糖基而进一步被修饰，因为糖基能促进其到达目标位置。

一种叫做信号识别颗粒（signal recognition particle，SRP）的 RNA 蛋白质

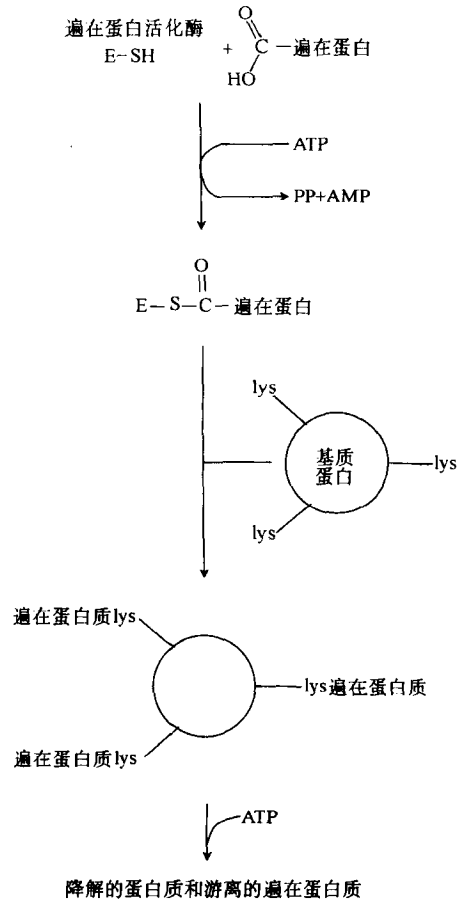


图 4-37　依赖于 ATP 和遍在蛋白的蛋白质降解过程

降解过程的第一步即遍在蛋白结合到基质上的过程需要 tRNA

复合体对蛋白质的移位过程起着重要作用，SRP 能与新合成的蛋白质的信号肽进行特殊互作。SRP 由长约 300nt 的 7SL RNA 和 6 种不同的、紧密结合在一起的蛋白质组成。SRP 能与核糖体结合，中止蛋白质合成。当多肽链增长到 70 个氨基酸时，即在信号肽完全通过核糖体时翻译过程被中止，此时，核糖体中还含有长约 30～40 个氨基酸的一段多肽链。SRP 并不能中止所有 mRNA 的翻译，只能特异性地作用于某些蛋白质如动物的分泌蛋白、植物的醇溶蛋白等的合成过程。当 SRP 与内质网膜上的信号受体蛋白接触时，mRNA 又恢复翻译过程。然后，SRP 扩散出去，启始另一条多肽链的运输过程。多肽链在移位

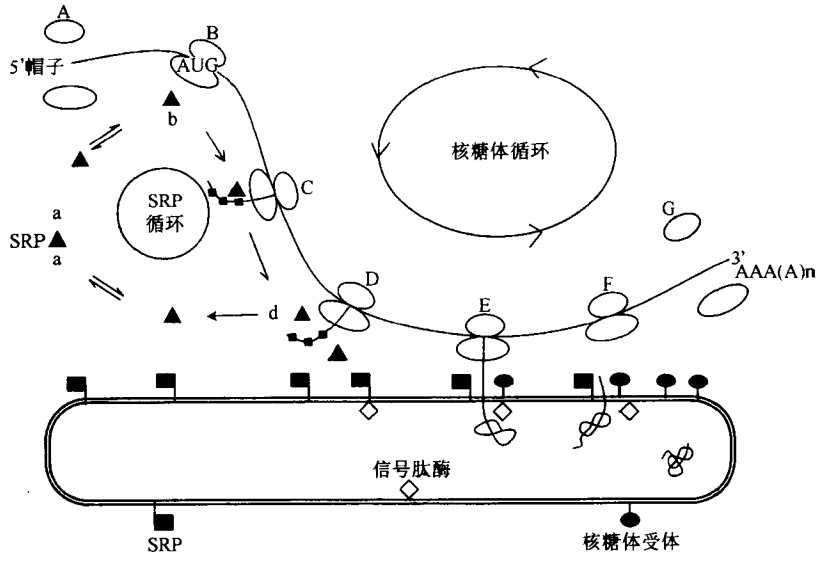


图 4－38　蛋白质穿越内质网的模型

A. SRP 首先与核糖体大亚基结合，当信号肽合成以后，SRP 阻止多肽链的延长（a－c）　B. SRP 与内质网上的 SRP 受体蛋白结合，脱离核糖体，使肽链继续延长（c－d）　C. 同时，核糖体大亚基与核糖体受体结合，内质网上的信号肽酶切割信号肽；然后 SRP 脱离 SRP 受体，进行新的循环（d－a）　D. 增长的多肽链通过某种未知的机制穿过蛋白质孔（E－F）　E. 多肽链的合成完成以后，蛋白质孔消失，核糖体进入细胞溶质进行新的循环

期间，聚核糖体附着到内质网膜的过程基本上由移位的多肽链来调节，但需要内质网中核糖体受体蛋白的帮助。

上述 SRP 对翻译过程的负调节作用可产生一种重要的生物学效应。例如，哺乳动物细胞分泌出去的许多蛋白质都是一些降解酶（如核酸酶和水解蛋白酶等）。即使这些蛋白质只在细胞内作短暂停留，也会对细胞本身产生很大危害。SRP 在中途阻止这类蛋白质的翻译，就能保证它们只有到达正确的膜位置时才能完整地合成。

现在除了知道蛋白质通过内质网移位的过程以外，还对蛋白质运输到线粒体、植物细胞的叶绿体（第五章）和细胞核的机制都有深入了解。由于这些细胞器具有双层被膜，所以蛋白质的移位过程就比由单膜组成的内质网更为复杂。

对于细胞核来说，许多核蛋白如组蛋白、hnRNP 蛋白、聚合酶以及有关因子等都是在细胞质中合成后进入核内的。现在发现，某些短的肽链具有将蛋白

质运输到核内的功能。可能某些蛋白质通过与将被运输到核内的蛋白质进行特异互作，而将其运输到核内。例如在滑爪蟾卵母细胞中有一种叫做核原生质配基（nucleoplasmin）的蛋白质，这种蛋白质具有将组蛋白运输到细胞核内并将其组装到 DNA 上的功能。

（六）超分子组装

基因表达在翻译后调节的最后一种方式是将新合成的有关蛋白质组装成功能单位。例如，血红蛋白和乳酸脱氢酶都由 4 条多肽链组成，二者的 4 个亚基都通过非共价结合而形成功能蛋白。还有一些蛋白质既可以多聚体形式存在，也可以非多聚体形式存在，如两种收缩性蛋白微管蛋白和肌动蛋白。对肌动蛋白而言，受精期间，它由球状单分子聚合成微纤丝的过程是动物精子顶体扩展过程的关键步骤。同样，在形成有丝分裂的纺缍体时，微管蛋白也组装成微管，只有在有丝分裂完成之后才变成单分子。这些单位微管蛋白然后又可组装成新的微管。

如前所述，还有几种蛋白质与核酸形成聚合体。以组蛋白为例，各种组蛋白只有与 DNA 结合，形成核小体才具有功能。又如，真核细胞的 80S 核糖体大约由 70 种蛋白质和 4 种不同 rRNA 组成，这些蛋白质和核酸只有结合在一起才具有功能。

各种超分子的组装都需要其他一些有关蛋白质的帮助，现在把这些参与各种超分子组装过程的蛋白质统称为侣伴蛋白（chaperon）（表 4－4）。侣伴蛋白是细胞内一些互不关联的蛋白质，它们只调节其他多肽链的正确组装，而自身又不是最终结构的组成成分。组装过程包括肽链折叠、寡聚化以及在蛋白质运输和行使功能期间折叠和寡聚化发生程度变化等。各种侣伴蛋白的基本功能可能是防止形成不具有生物学功能的不正确结构，抑制多肽链不同部分之间以及不同分子之间不正确的互作，包括相同或不同多肽链、各种核酸、甚至一些小分子代谢产物之间或分子内的不正确互作。虽然有些分子能够通过自动组装形成聚合体，但这种自动组装过程可能仍然需要有关侣伴蛋白的帮助。侣伴蛋白还可限制由环境压（如高温）所造成的损坏，因此可能各种热激蛋白都属于侣伴蛋白的范畴。侣伴蛋白也与核小体和植物 Rubisco 酶的组装有关。它在组装的早期阶段识别这些蛋白质的结构特征并同其结合，从而阻止不正确的组装过程。值得指出的是，在各种细胞、各种生物的线粒体以及植物叶绿体中都有一组叫做侣伴蛋白宁（chaperonin）的侣伴蛋白，在各种生物中，其氨基酸序列都非常相似，而且是一类含量丰富的组成型蛋白。当细胞受热激或细胞内未折叠的蛋白质含量增加时，其含量也随之增加，因此各种侣伴蛋白宁参与调节各种

细菌、线粒体和质体中许多蛋白质的组装过程。

表 4－4　侣伴蛋白家族

名　　称	功　　能
核原生质配基	核小体组装、转录？核糖体组装和运输？
各种侣伴蛋白宁	蛋白质折叠、寡聚体组装，蛋白质运输、DNA 复制，mRNA 更新、阻止环境压
热激蛋白 70 类	蛋白质折叠、寡聚体组装、蛋白质运输、寡聚体拆除？
热激蛋白 90 类	掩饰激素受体
信号识别颗粒	蛋白质运输
枯草杆菌蛋白酶的前序列	枯草杆菌蛋白酶的折叠
α－裂解蛋白酶的前序列	α－裂解蛋白酶折叠
遍在蛋白化的核糖体蛋白	真核细胞核糖体组装
发射因子	蛋白质运输
SccB 蛋白	蛋白质运输
PapD 蛋白	大肠杆菌性纤毛组装

第五节　发育阶段对基因表达的调节

一、基因表达的细胞、组织和器官特异性

多细胞生物的发育都是从一个受精卵开始的。受精卵经过一系列细胞分裂和分化，形成各种特化组织，以执行特殊功能。一个成熟的动物大约有 200 种不同的细胞型，尽管每一类型的细胞的基因组基本相同，但它们表达不同的基因。例如，组成血红蛋白的 α－和 β－珠蛋白主要在红血细胞中合成，胰岛素只在胰腺中合成，而生长素则主要在脑垂体内合成。又如，植物在发育期间，某些特殊蛋白质如种子贮藏蛋白、叶绿体蛋白、与色素形成和果实中贮存物合成有关的酶类都只局限在某些特殊结构或器官内。因此，有关基因只在某些特殊细胞、组织或器官中表达。

哺乳动物基因组中脱辅基脂蛋白（apolipo protein－B，Apo－B）基因是研究较清楚的具有组织或器官特异性表达特点的基因之一。Apo－B 蛋白是血清中一类载脂蛋白，对哺乳动物的动脉粥样硬化（athrosclerosis）起主要作用。哺乳动物基因组只含有一个拷贝的 apo－B 基因，该基因有 29 个外显子，整个编码区由 4563 个密码组成。在肝细胞中，其 mRNA 的整个编码区都可以编码，翻译出 512kd 的蛋白质。但是在肠细胞中，Apo－B 基因只编码出一种较短的、约 250kd 的蛋白质。肝细胞和肠细胞中编码这两种蛋白质的 mRNA 只有一个碱基差别，即第 2153 位的密码子中碱基 C 变成了 U。这种转换使编码谷氨酰胺的密码子 CAA 变成终止密码 UAA（图 4－39）。一般认为，造成这种转换的原因是 mRNA 在转录后通过 RNA 编辑而实现的。

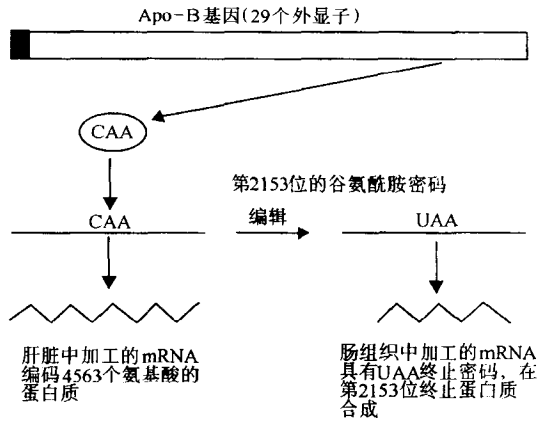


图 4－39　apo－B 基因的组织特异性表达

通过 RNA 编辑控制基因的组织特异性表达的另一例子为小鼠的脑组织特异性表达的谷氨酰胺受体基因（glutamate receptor GluR）。该基因在转录后只将其中一个谷氨酸密码 CAG 转变成 mRNA 分子中编码精氨酸的密码 CGG。精氨酸在控制离子通过神经递质的流动过程中起着重要作用。显然，这种 RNA 编辑事件对神经组织执行特殊生理功能是必需的。

apo－B 基因的 RNA 编辑事件表现为碱基 C 到 U 的转换，而谷氨酰胺受体基因的 RNA 编辑则表现为碱基 A 到 G 的转换。这两种转换都是通过脱氨基作用即将核苷酸环上的氨基除掉而实现的，二者分别由胞嘧啶核苷脱氨基酶和腺嘌呤核苷脱氨基酶所催化。虽然这种类型的 RNA 编辑与脱氨基酶有关，但编辑反应是在一种复合体中进行的。这种复合体除了含有一个具催化作用的脱氨基酶以外，还含有其他亚基。额外亚基可能帮助编辑机器识别特殊的编辑位

点，不同细胞型或不同组织可能决定编辑机器中有关辅助亚基的使用。

二、脊椎动物珠蛋白基因表达的发育阶段调节

有关发育和细胞分化调节基因表达的最熟知的例子是红血细胞的珠蛋白基因。具有功能的珠蛋白基因在所有脊椎动物中都相同，各个基因的编码序列都有3个外显子，可能它们都起源于相同的祖先基因。

在成体细胞中，珠蛋白四聚体由两条相同的α链和两条相同的β链组成。但是在许多动物包括鸡和人体的胚胎或胎儿细胞中的血红蛋白都与成熟个体红血细胞中的血红蛋白不同（图4－40）。在人体中，α链类包括ζ（Zeta）和α链，β链类则包括ε、γ、δ和β链。ζ链是最先合成的类α链，但不久后就由α链本身所取代。ε是最先合成的类β链，后来就由δ和β链所取代。人体胚胎血红蛋白主要由2条ζ珠蛋白链和2条ε珠蛋白链和4条血红素分子组成。在怀孕后的第二个月，ζ和ε珠蛋白合成骤然停止，而α和γ珠蛋白的合成则增强。在怀孕后第3个月，β和δ珠蛋白基因开始活动，随着α珠蛋白合成量逐渐降低，这两种珠蛋白的含量慢慢增加，而且这种转变速率在出生后明显加强，从而胎儿血红蛋白就被成体血红蛋白 $\alpha_2\beta_2$ 所替代。在正常的成熟个体中，$\alpha_2\beta_2$ 约占97%，$\alpha_2\delta_2$ 约占2%～3%，$\alpha_2\gamma_2$ 约占1%。

在所有高等灵长目动物中，珠蛋白基因家族包括两个基因簇，即β链类基因簇和α链类基因簇（图4－40）。β基因簇长约50kb，含有5个功能基因（即ε基因、2个γ基因、δ基因和β基因）和一个假基因。两个γ基因 γ^A 和 γ^G 只有一个氨基酸的差别，前者在肽链的第136位为丙氨酸，后者为甘氨酸。α基因簇长约28kb，包含一个ζ功能基因，一个ζ假基因、2个α功能基因，2个α假基因和一个未知功能的θ基因。

在人体基因组中，β珠蛋白基因簇位于第11染色体上，α基因簇位于第16染色体上，因此二者必须协调表达才能等量合成两种多肽。染色体11上的β基因簇的各个基因的表达次序是依次从胚胎珠蛋白转变到胎儿和成体的珠蛋白。

β珠蛋白基因簇在发育过程中的调节表达向我们提出了两个主要问题，即为什么这些基因只在分化的红血细胞中表达？它们的表达是如何从基因簇的一个成员转变到另一个成员的？当将一个额外的、成熟小白鼠的β珠蛋白基因引入一个小白鼠胚胎基因组中后，该基因只在红血细胞中表达，并且能够在发育期间进行适当的发育转换。这就表明细胞类型和发育阶段对基因表达起着控制作用，但目前对转换信号及其特征还不清楚，它可能与γ和δ基因的一段重复序列有关。另外，当将绵羊胚胎的红血细胞移植到成熟动物体内时，这些红血

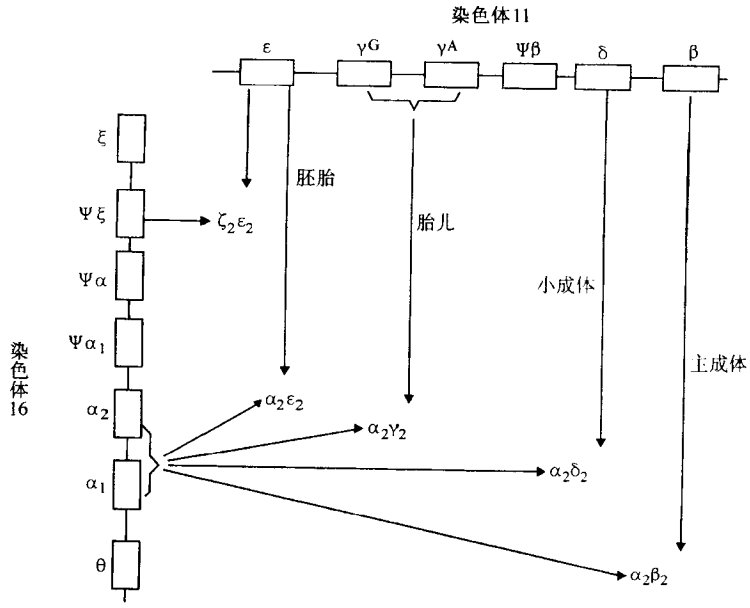


图 4－40　人体发育期间血红蛋白基因活动的先后顺序

细胞能够连续几天转录胚胎珠蛋白基因，然后才转换成转录成体的β珠蛋白基因。这种延迟现象表明，成体动物中基因表达的转换不是由于某种细胞外信号引起的，而是红血细胞遵循一种发育转换的细胞内程序。

在各种基因家族中，基因表达从一种形态转变到另一种形态的例子除了β珠蛋白基因簇以外，还有其他一些基因家族如组蛋白基因、肌动蛋白基因、微管蛋白基因以及昆虫的卵壳蛋白基因等。

三、性别对基因表达的调节

（一）果蝇性别对基因表达的调节

果蝇的性别由X染色体和常染色体共同决定，其X染色体上的雌性定子的平衡性决定雌性发育，常染色体上的雄性定子的平衡性决定雄性发育。Y染色体对性别决定不起任何作用，它只在发育的后期精子形成时才起作用，只与雄性的可育性有关。如果二倍体细胞中只有一条X染色体（1X:2A），该个体为雄蝇；如果为两条X染色体（X:A:1），则表现为雌性。

果蝇性别表现型受一组基因控制，这组基因或者将X:A比例转变成雄蝇比例，或者转变成雌蝇比例（图4－41）。这一途径的关键基因是转化基因

(transformer，tra)。转化基因是产生雌蝇所必需的。如果没有这一基因，无论X:A比例如何，都会产生雄蝇。在整个幼虫阶段，转换基因积极转录，其初级转录子或者被加工成一种通用mRNA（即在雌、雄蝇中都存在），或者加工成雌蝇特异性的mRNA（图4－42）。只有雌蝇含有可变剪接的mRNA，它可能是这一基因的惟一具有功能的mRNA。通用mRNA在第二个外显子上游有一个终止密码，这种mRNA指导合成一种不具功能的小蛋白质，因此不具有决定性别的作用。但是在雌蝇特异性的mRNA中，这一UGA终止密码位于内含子中，它可以在mRNA成熟过程中被除掉，所以并不影响mRNA的翻译。如果将这一雌蝇特异性转录子的cDNA整合到XY果蝇基因组中，可使XY果蝇成为雌蝇。

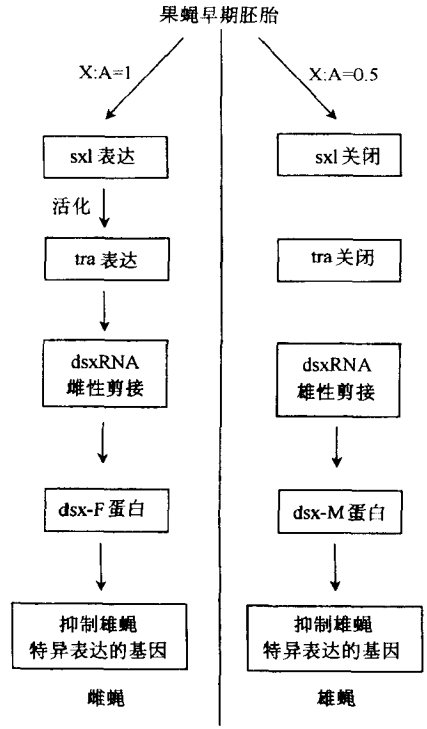


图4－41　果蝇性别决定示意图

示性别决定基因活动顺序与性别表型发育，如果性致死基因被活化，转化基因也被活化，从而合成雌蝇特异性的mRNA；当存在雌蝇特异性的转换基因产物时，两性基因的转录子以雌蝇特异性方式加工，产生雌蝇表型；如果转换基因不能合成雌蝇特异性产物即性致死基因未被活化，两性基因的转录子则以雄蝇特异性方式剪接，产生雄蝇表型

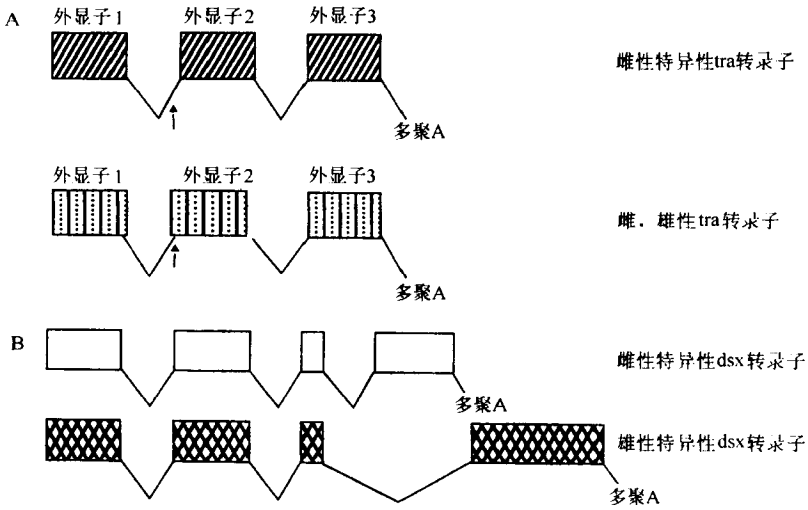


图 4－42　果蝇发育期间性别特异性转录子的合成策略

A. 由转化基因编码的雌蝇特异性的和通用转录产物通过可变 RNA 剪接而形成。箭头示 UGA 终止密码，这一密码只存在于通用转录子的外显子中，而不存在于雌蝇特异性的成熟 mRNA 中。方块示外显子，细线示内含子　B. 两性基因的可变 RNA 剪接。化蛹期间，依据 tra 和 tra2 基因是否被活化，两性基因将合成雄蝇特异性的或雌蝇特异性的产物

是什么因素使转换基因只在 XX 细胞中而不在 XY 细胞中加工出雌蝇特异性的转录子呢？可能这种性别特异性的可变剪接与内含子中 2 个 3′剪接位点的竞争有关。研究表明，这种竞争受一个性致死基因（sex lethal gene，sxl）调节。性致死基因是果蝇性别决定途径中最早表达的基因之一，它在转化基因之前起作用。如果 X:A 比例为 1，性致死基因首先被活化。当 sxl 基因活化后，tra 基因就合成通用的或雌蝇特异性的转录子，胚胎就发育成雌蝇。如果不存在 sxl 基因或 sxl 基因发生突变，tra 基因只合成不具功能的通用转录子，胚胎则发育成雄蝇。通常认为，sel 基因控制 tra 基因的转录子使用哪一个 3′剪接位点。

性致死基因采用两种主要方式控制剪接位点的使用（图 4－43）。其方式之一是阻止使用通用受体位点而允许使用可变的、雌蝇特异性的受体位点。另一种方式是活化雌蝇特异性剪接位点。支持第一模型的证据来自利用 P 转座子的转化实验。当将一个含有不具功能的通用 3′受体位点的基因通过 P 转座子插入到 XY 胚胎的基因组中以后，转化的胚胎开始合成雌蝇特异性转录子，并发育成雌蝇。后来发现，由 sxl 基因编码的产物蛋白质直接结合到 tra 基因的前体 mRNA 上，这一结合位点在非特异性受体位点附近。如果 tra 基因的通用

3′剪接位点由于突变或被 sxl 基因产物所阻碍，在 mRNA 的加工过程中则使用可变的、雌蝇特异性受体位点，其结果是产生雌蝇。

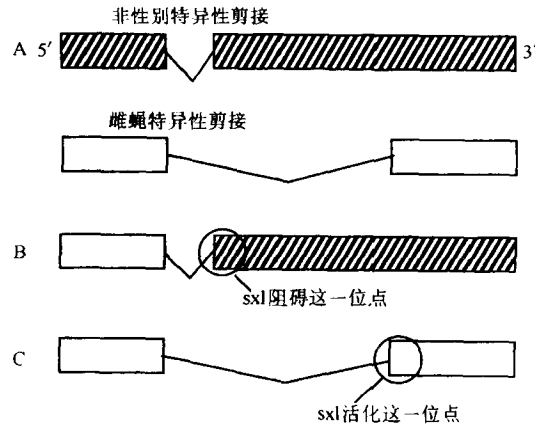


图 4－43　性致死基因产物调节 tra 基因转录子加工的两种主要途径

在果蝇的性别决定途径中，一般认为 tra 基因控制所谓两性基因（double sex gene，dsx）的表达。dsx 基因是形成两种性别时都需要的。dsx 基因中发生突变，可使 XX 胚胎发育成雄蝇，或使 XY 胚胎发育成雌蝇。化蛹期间，dsx 基因转录出一种以两种可变方式进行加工的转录子，或者产生雌蝇特异性 mRNA，或者产生雄蝇特异性 mRNA（图 4－43B）。在雌蝇和雄蝇中，两种 mRNA 开头的 3 个外显子相同，只是第 4 个外显子不同。雄蝇特异性的 mRNA 缺失了一大段，其中包括雌蝇特异性外显子。

转化基因的作用发生在两性基因之前，合成一种剪接因子，使 dsx 基因的前体转录子加工成雌蝇特异性的 mRNA。在这种情况下，tra 基因产物为一种剪接因子，类似性决定致死基因的产物。如果 tra 基因发生突变或未被活化（如 XY 果蝇）那么两性基因的转录子就以雄蝇特异性方式进行加工，产生雄蝇。可变 mRNA 加工在果蝇的整个发育阶段起着极其重要的作用，它控制发育阶段的转变，同时也是不同发育阶段调节同一基因进行差别表达的主要方式之一。

（二）基因组刻印

基因组刻印（genome imprinting）现象是指基因组中一小组基因依据其亲本来源，在胚胎发育或个体发育期间进行单等位基因表达的现象。当某个被刻印座位的一个等位基因通过一种性别的配子发生而传递时，该基因的表达活性不受影响；但是当这一基因通过另一种性别的配子传递时，其活性受到抑制而

不能表达。因此就把这一失活的基因叫做被刻印基因（imprinted gene）。例如小鼠的编码类胰岛素生长因子（insulin－likegrowth factorⅡ，IGF－2）的等位基因（位于小鼠的第7染色体上）只是其中一个从父本遗传的等位基因表达，而从母体遗传的一个等位基因则不表达。而位于小鼠第17染色体上的，编码IGF－Ⅱ受体的基因Igf－2r则只是其中一个源于母本的等位基因表达。

对于基因组刻印现象目前有几种解释，其中一种观点是被刻印基因的亲本特异性表达依赖于母本－父本基因组之间或两个等位基因之间的互作。互作过程导致1个或2个等位基因发生差别基因表达（differential gene expression）。在果蝇中，这种现象叫做转位（transvection）或显性位置效应花斑（dominant position effect variegation），在玉米和马铃薯中则叫做副突变（paramutation）。果蝇中显性位置效应花斑现象与反式等位基因失活（被刻印基因）有关，而失活过程则由异染色质化所调节，即通过染色体重排，将一个野生型的常染色质基因并列到同源染色体的异染色质区附近而使其失活。另一方面在某些细胞当中，当异染色质结构扩展到所谓的花斑座位中以后，也可使常染色质基因发生转录失活。异染色质的扩展可能受某些显性修饰因子控制，这些修饰因子或者促进异染色质形成，或者抑制异染色质形成。

另一种观点认为基因组刻印现象与基因组互作无关，它可能是配子形成期间，DNA发生特异性甲基化的结果。例如在小鼠胚胎发育的早期，父本起源的Igf－2等位基因不甲基化，能够表达，而母本来源的一个等位基因则被甲基化，因此不能表达。在配子形成过程中，如果为雄鼠，进入精子中的一个等位基因不管原来是否甲基化，必然是未甲基化的，因此处于精子中的一个母本等位基因必须脱甲基化。如果为雌鼠，进入卵子中的一个等位基因必须甲基化，所以如果该基因原来为父本等位基因，此时必须加上甲基。但是，在目前已知的被刻印基因中，大约只有70％的基因发生甲基化，表明甲基化作用可能只是造成被刻印基因的亲本特异性表达的机制之一。

第三种假说认为，基因组刻印是受基因控制的，可能涉及数种基因产物。这些基因叫做所谓的刻印基因（imprinting gene），它们协调表达，最后产生一种蛋白质复合体，从而影响或稳定染色质结构。

第六节　环境信号对基因表达的调节

一、半乳糖对酵母基因表达的调节

酵母细胞为了利用半乳糖进行糖酵解，必须活化4个基因gal1、gal2、

gal7 和 gal10。gal2 基因产物的功能是将半乳糖运输到细胞中，gal1、gal7 和 gal10 是半乳糖转变成葡萄糖－1－磷酸脂所必需的。如果培养基中加入半乳糖，可使细胞中这些基因的转录水平提高 5 000 倍以上。另一方面，葡萄糖可以抑制各个 gal 基因的诱导表达过程，因此当培养基中存在葡萄糖时，这 4 个基因都被抑制。这 4 个决定半乳糖代谢的基因受两个调节基因 gal4 和 gal80 的控制。这 6 个基因通过互作调节酵母半乳糖代谢过程（图 4－44）。

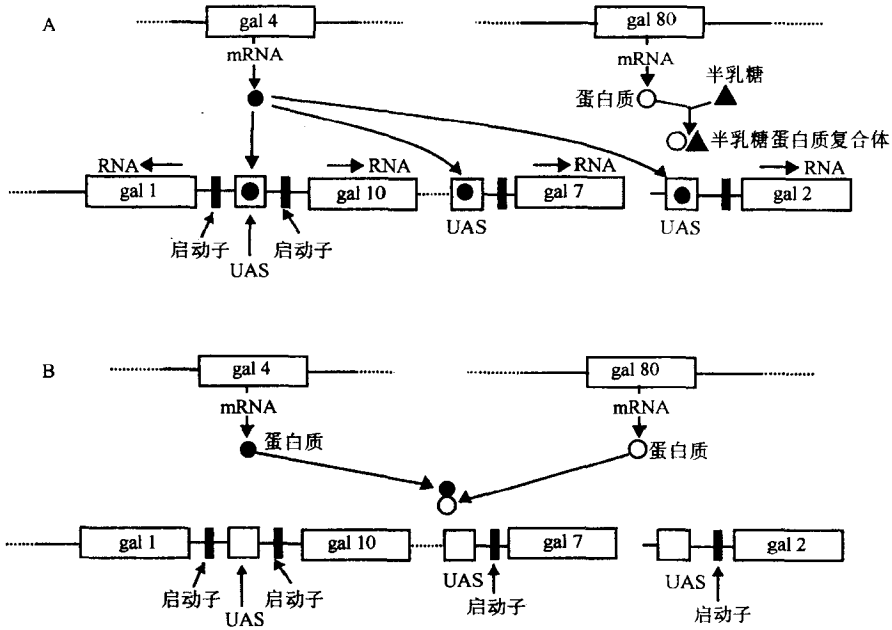


图 4－44　酵母半乳糖代谢基因表达的调节模型

A. gal4 基因编码的蛋白质通过与 gal1、gal2、gal7、gal10 基因的 UAS（上游活化位点）结合，活化这 4 个基因，gal1 和 gal10 共用二者之间的一个 UAS，二者以相反的方向转录，gal80 蛋白与半乳糖结合后，就不能与 gal4 蛋白互作　B. 当培养基中不存在半乳糖时，gal80 蛋白与 gal4 蛋白互作，阻止 gal4 蛋白活化 4 个半乳糖代谢基因的转录

gal4 基因编码的产物是 4 个半乳糖代谢基因 gal1、gal2、gal7、gal10 的表达所必需的。gal4 基因为组成型表达，其编码产物通过与各个基因的转录起点上游长约 200bp 的序列结合，活化 4 个半乳糖代谢基因的转录。gal1、gal7 和 gal10 三个基因位于同一染色体上，它们都分别进行转录。其他 gal 基因则分布在其它不同染色体上。gal4 蛋白的实际结合位点由长 17bp 的 4 段顺接重复序列组成，这种序列叫做上游活化位点（upstream activationsite，UAS）。上游活化位点与启动子序列如 TATA 框等一起，决定基因的转录启始。

gal80 也是一个组成型表达的调节基因。在不存在半乳糖时，其编码产物蛋白质阻止 gal4 蛋白与 4 个基因的 UAS 序列结合。如果存在半乳糖，半乳糖则与 gal80 蛋白结合，阻止 gal80 蛋白抑制 gal4 蛋白的活性，其结果是诱导编码半乳糖代谢酶的基因转录。葡萄糖可以直接或间接地与 gal4 蛋白结合，阻止 gal4 蛋白与 UAS 序列结合，从而抑制转录诱导过程。

半乳糖除了调节上述 6 个基因的转录以外，还可调节酵母其他基因的转录。

二、外源激素对基因表达的调节

（一）类固醇激素对基因表达的正调节

类固醇激素（steroid hormon）对各种动物乃至人体的发育都具有重要调节作用。用类固醇处理敏感细胞，通常只在几分钟内就能刺激合成一些特异性的、新的 RNA。类固醇与叫做类固醇激素受体（steroid hormon receptor）的蛋白质结合后，这种激素/受体复合体然后再与基因组中特殊座位互作，启始有关基因转录。因此类固醇的主要作用是在转录水平上调节基因表达。

最为熟知的例子之一是糖皮质激素（glucocorticoid）对基因表达的调节。这种激素的受体蛋白的分子量很大，由 4 个相同的、分子量为 94kD 的亚基组成。当受体与一种人工合成的糖皮质激素地塞米松（dexamethasone）结合后，这种复合体结合到某些特殊 DNA 序列即所谓的依赖于激素的增强子上，活化转录启始过程。所有受类固醇调节的基因中都存在激素/受体复合体的结合位点，但在离体条件下，激素/受体复合体的结合过程有时不受这些位点的影响。利用 DNA 重组技术，将这些激素/受体的结合序列与不受激素调节的基因的启动子融合后，糖皮质激素也能刺激这种融合基因在活体内表达。

受激素活化的基因在受到激素的刺激后，其所在区域染色质发生几种结构变化。第一，在加入地塞米松后，染色质对 DNA 酶 I 的总体敏感性增加，并且在激素消除后很久仍维持这种敏感性。第二，在激素/受体结合位点附近产生不同的 DNA 酶 I 过敏性区，当激素消除后，这些过敏性区也随之消失。有些基因的表达还受一种以上类固醇激素调节，如糖皮质激素和黄体酮都能调节鸡溶菌酶基因表达。这类基因具有几个不同的与不同激素/受体复合体结合的位点。所有这些现象表明，各种类固醇激素的受体蛋白可能通过与类似增强子的序列互作，活化有关基因转录。

类固醇还可通过其他机制调节基因表达。例如，它可以有选择地稳定有关转录产物，阻止其在细胞内降解（表 4－5）。在滑爪蟾肝细胞中，在存在诱导激素雌激素的条件下，卵黄前体蛋白（vitellogenin）mRNA 的半衰期约为三

周，当消除雌激素后，半衰期下降到16小时。另一方面，当用类固醇激素处理敏感细胞之后，往往会改变RNA的加工过程，在某些情况下，还可提高总体RNA的合成速率，从而导致增生生长。总而言之，类固醇激素通过调节基因转录和转录后过程而使某些特异mRNA大量累积，提高基因的表达水平。

表4－5　由激素稳定的某些特殊mRNA

（根据Shapiro等，1987）

mRNA	细胞或组织	调节效应子	半衰期（小时）	
			有效应子	无效应子
卵黄前体蛋白	非洲蟾蜍肝脏	雌激素	500	16
清蛋白	非洲蟾蜍肝脏	雌激素	10	3
卵黄前体蛋白	禽肝脏	雌激素	22	~2.5
Apo VLDL Ⅱ	禽肝脏	雌激素	26	3
酪朊	小鼠乳腺	催乳激素	92	5
生长激素	小鼠脑下垂体细胞培养物	地塞米松和甲状腺素	20	2
胰岛素	小鼠胰岛细胞	葡萄糖	77	29
卵清蛋白	母鸡输卵管	雌激素和黄体酮	~24	2~5

（二）植物生长素对植物基因表达的调节

植物生长素（auxin）是调节植物生长发育的主要激素之一，同动物类固醇相似，其基本功能是在基因的转录水平上调节某些特殊mRNA的合成。例如豌豆上胚轴（epicotyl）在用IAA处理后，能够合成6种特殊的多聚A化的mRNA，其中有3种mRNA的翻译产物出现在处理后20min，1种出现在处理后30min，2种出现在处理2h以后。利用IAA、2，4－D等类生长素处理其他植物的组织也都得到相似结果。

植物生长素除了诱导植物合成某些特殊mRNA以外，还可抑制某些mRNA合成，这种生长素对基因转录的负调节作用与所使用的浓度呈正相关。也就是说，生长素浓度愈高，抑制效率愈强，反之则愈弱。

三、热激效应对基因表达的调节

当温度徒然升高时，所有生物的细胞都会改变其基因表达的方式，合成某些特殊蛋白质如热激蛋白，同时停止合成大多数多肽。根据对果蝇的研究，当温度升高到37℃时，果蝇多线染色体在一两分钟内就可以在几个区域出现疏松，即所谓热激座位。当热激消退后，多线染色体则在其他一些区域出现疏松。疏松现象反映了转录机器从正常状态转变到对热激基因进行特异性转录的

状态，同时也反映了细胞内的翻译机器也发生相应变化，以保证热激 mRNA 进行倾向性翻译。

在高等植物中，热激蛋白大致可分成三种不同类型，第一组为大分子量蛋白，其分子量变动于 68～104kd 之间，这组蛋白质普遍存在于所有生物之中，包括细菌、动物和植物。第二组为中等大小的分子，其分子量在 20～30kd 之间。第三组为小分子蛋白，其分子量在 15～18kd 之间，为高等植物所特有。当植物处于高温压（40～45℃）下时，同时合成上述三组蛋白质。

同果蝇一样，热激对植物的遗传效应主要表现为改变基因的转录过程。例如，大豆在受热激处理 2h 内，合成 20 种不同的热激蛋白 18 的 mRNA，其拷贝数在每一细胞中可高达 2 万。当受热激处理的组织回到正常温度下时，数小时后细胞中热激基因转录子的含量以及热激蛋白的合成速率均迅速下降。

热激也可以在翻译水平上抑制正常蛋白质的合成。例如果蝇在热激处理期间，虽然许多正常蛋白质的合成受到抑制，细胞中只合成热激蛋白，但是各种正常 mRNA 并不降解，只是处于不活动状态。当温度恢复正常、在未形成新的正常转录子之前，细胞照样合成正常蛋白。

在各种真核生物中，热激反应基因通常为多基因系。例如果蝇基因组含有 7 个热激基因，这些基因的转录活化都受启动子中某些特殊序列的控制。对其中一个热激基因 hsp70 来说，在 TATA 框的上游有 4 个转录因子的结合位点（图 4－45），其同感序列为 C－－GAA－－TAA－－G。这种热激控制信号序列具有回文对称性，而且在 7 个热激基因中都高度同源。

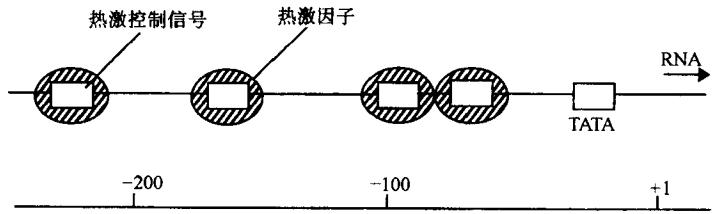


图 4－45　果蝇 hsp70 基因上游热激因子的 4 个结合位点

各种真核生物的热激基因在结构和功能上也有很大相似性。如玉米热激基因 hsp70 的编码区与果蝇同一基因的相应部分存在 75% 的序列同源性。又如果蝇的热激信号序列也见于大豆的热激基因之中，有所不同的是，大豆热激基因在该位点（－48～－62）上游还有另外 2 个次级同感序列 AGAAATTTTCT 和 CTNGAANNTTCNAG。这两种次级结构保守性序列可能对大豆热激基因的表达也具有重要调节功能。由于所有热激信号序列都具有回文对称性，所以它们都是一些 DNA 结合蛋白的识别位点。但目前对热激作用如何调节热激因子本身的活性还不清楚。

四、光照对植物基因表达的调节

（一）光受体与基因表达

植物在进行光合作用过程中，首先就是光受体（photoreceptor）吸收光照，然后再将光信号传导到某些特殊基因，改变这些基因的表达状况。目前研究较为深入的光受体主要有三种，即原脱植基叶绿素（protochlorophyllide）、光敏色素（phytochroine）和蓝光受体（blue－light receptor），其中尤以光敏色素研究最为深入。光敏色素为一种吸收红光的色素，以两种形态存在，即非活性形式（Pr）和活性形式（P_{fr}）。光敏色素吸收红光或远红光后，可以从非活性形式转变成活性形式，或者相反。

$$Pr \underset{\text{远红光感应}}{\overset{\text{红　光}}{\rightleftharpoons}} P_{fr} \xrightarrow{\text{传导}} \text{改变基因表达} \xrightarrow{\text{反应}} \text{形态发生}$$

根据上述光敏色素的转换情况和光信号的传导方向，光敏色素起着一种双重分子开关的作用，对来自环境的光信号作出反应，调节植物基因表达，最后改变植物生长发育方式，使其适应光照环境。

对绿藻的研究发现，植物细胞至少含有2种光敏色素即白化组织光敏色素（第一类）和绿色组织光敏色素（第二类）。第一类光敏色素又叫光敏色素A，大量存在于白化组织之中，是细胞中迅速降解的P_{fr}。第二类光敏色素在白化组织中含量很低，为一种稳定的P_{fr}。

所有光敏色素的初级结构都基本相同，长约1 110～1 172个氨基酸（124～129kD），其NH_2半部分都有一个生色团（chromophere）结合位点。光敏色素A通过折叠形成的次级结构可分成2个主要部分，即分子量为70kD的球状NH_2－末端部分和分子量为55kD的长形COOH－末端部分。这两部分结合在一起形成一个非对称的二聚体。

（二）基因对光反应的特征

光照对植物的某些细胞核基因和质体基因表达都具有调节作用。如由光敏色素控制的细胞核基因有1，5－二磷酸核酮糖羧化酶（Rubisco）小亚基基因（rbcS）、集光叶绿素－蛋白质复合体的叶绿素a/b蛋白基因（cab）以及原脱植基叶绿素还原酶基因和光敏色素基因自身等。但是光敏色素对后二者的调节呈负调节反应，即光敏色素降低这两个基因的转录速率。在黑暗条件下，原脱植基叶绿素还原酶mRNA和光敏色素mRNA的含量在细胞中累积很高，当光诱导形成P_{fr}时，其mRNA含量又迅速下降。

在所有由光诱导表达的基因中，不同基因对相同光源反应各不相同，例如用红光照射豌豆幼苗后，其组织中累积有 11 种由光敏色素调节的转录子，呈 4 种累积形式，其中有 6 种呈直线式累积，2 种迅速增加后再保持平稳，2 种在后期受到抑制，一种通过红光闪烁可增加累积速率。cab 基因和 rbcS 基因的转录子属第一种，而铁氧还原蛋白（ferredoxin）基因的转录子则属第二种类型。

同一基因家族中不同成员对光照的反应也有差异。例如用白炽光处理豌豆白化组织 24 小时后，rbcS 基因家族的两个成员（rbcS－3A 和－3C）的 mRNA 累积量很高，而另外两个成员（rbcS－E9 和－8.0）则几乎不受影响。这种差别与叶片的发育阶段也有关。在成熟的绿叶中，rbcS－3A、3C 和 8.0 的 mRNA分别占整个转录子含量的 40%、34%和 19%，而 rbcS－E9 和另外两个 rbcS 基因的转录子总共才占 7%左右。

不同质量的光源对调节基因表达具有不同效率。例如在未成熟的豌豆组织中，红光能够诱导 rbcS 基因表达，但是在成熟叶组织中蓝光则比红光的诱导效率更高。此外，光照的强度也可影响光调节基因的表达。

根据对不同植物 rbcS 基因家族中不同成员的研究，这些基因在转录起点上游都有一至数段与接受光信号有关的 DNA 序列，这类序列叫做光调节子（light regulatory element）。它们具有与 SV40 增效子相似的 GT 序列，其同感序列为：

G　　TT　T　T
TGTGG　A　TA　G
A　　CC　A　A

细胞中某些蛋白质通过与其它因子竞争，与光调节子结合。

在浮萍（*Lemna*）中，红光可诱导一种核因子（LRF－蛋白）与 rbcS 基因启动子中光调节子互作的程度发生改变，这种蛋白质可能就是光信号传导途径的一部分，它调节由光敏色素控制的基因的转录。除了细胞核因子以外，某些质体因子也调节核基因的光调节转录，如在某些叶绿体缺失突变体中，光照就不能诱导 cab 基因转录。

光照除了在转录水平上调节某些基因表达以外，而且还可以在翻译水平上调节这些基因表达。例如将在光照条件下生长的苋菜（*Amaranthus*）幼苗移到黑暗条件下栽培，虽然 Rubisco 的大小亚基 mRNA 都能与核糖体结合，但是大、小亚基蛋白都不能正常合成。

五、缺氧对植物基因表达的调节

植物通过缺氧处理会使其正常蛋白质的合成受到抑制而合成某些特殊多

肽。通常植物在遭洪水淹没时都会作出这种反应。例如玉米幼苗在遭水淹没后，在开始的5h内合成一种分子量为33kD的多肽，叫做过渡性多肽（transitional polypeptide），大约90min后开始合成20种额外多肽，这组缺氧多肽约占细胞中总体蛋白质含量的70%以上。然后过渡性多肽停止合成，但其含量在整个缺氧期间都一直保持很高水平，而缺氧多肽则以恒定速率继续合成，直到缺氧处理72h后，细胞死亡时其合成速率才开始下降。

玉米中这组缺氧多肽主要是由两个酒精脱氢酶基因Adh－1和Adh－2编码的酒精脱氢酶的同功酶，此外还包括葡萄糖－6－磷酸异构酶、果糖－1，6－二磷酸醛缩酶、蔗糖合成酶等。由于玉米Adh－1基因的缺失突变体在淹水5h后就死亡，所以植物缺氧反应的主要功能之一是尽可能地生产ATP，以满足细胞能量代谢的需要。

缺氧条件直接影响基因的转录过程。玉米在缺氧处理期间，各种缺氧特异性mRNA急剧增加。以Adh－1和Adh－2基因产物为例，缺氧处理90分钟后，其特异mRNA开始增加，其中Adh－1基因的mRNA直到48小时后仍保持平稳水平，Adh－2基因的mRNA在10小时后开始下降。这些特异mRNA含量的升降过程与其蛋白质合成的变化非常相似。目前对缺氧条件如何作用于基因转录过程的机制还不清楚，可能同热激效应相似。缺氧条件也可能促进某种特殊转录调节因子与基因启动子中某些特殊序列互作，提高有关基因的转录效率。

小　结

真核生物的DNA与4种基本的组蛋白结合，形成叫做核小体的结构。核小体的核心是组蛋白H_2A、H_2b、H_3和H_4各以2个分子组成的8聚体。DNA以140对碱基的序列绕8聚体2圈，另一种组蛋白H_1结合到核小体外的DNA上，形成核小体重复。核小体重复是染色质的一级结构。染色质进一步螺旋化，形成直径为30nm的超螺旋体或光学显微镜下可见到的染色体。真核细胞中的染色质有两类，一类为螺旋化程度高的、不具转录活性的异染色质，另一类为松弛的、具转录活性的常染色质。

真核细胞的每一条染色体都有一个着丝粒，两端各有一个端粒。着丝粒和端粒都由高度重复的DNA组成，在基因组中形成衔接重复；通过氯化铯梯度离心，以所谓的卫星DNA带形式出现。高等真核细胞中还存在分散的高度重复DNA。人体基因组大约含有10万个拷贝的、分散的Alu序列，每一拷贝的中部都有一个由限制性内切酶AluI切割的位点。细胞分裂中期的染色体是由

许多染色质环附着到一种蛋白质骨架上形成的。

真核生物的每一条染色体都有多个复制启始区即含有多个复制子。不同物种的DNA复制可能都涉及多种不同的DNA聚合酶和有关辅助蛋白质。DNA复制发生在细胞周期的S期，其复制过程基本与大肠杆菌DNA复制相似，不同的是采用两种不同DNA聚合酶分别催化前导链和后随链的延长。细胞中普通复制机器不能复制染色体两端的端粒，这种结构由端粒酶以其自身RNA为模板，按5'→3'方向延长前导链即G链，另一条互补链合成则由正常复制机器完成。

真核生物在一定的发育阶段可以有选择地扩增某些基因。细胞内的基因扩增有三种不同形式，第一种是滑爪蟾rRNA基因在卵子发生期间进行扩增，一个rRNA转录单位首先脱离染色体，然后按滚环模型进行复制。第二种类型以果蝇卵壳蛋白基因扩增为代表。卵壳蛋白基因扩增时不脱离染色体，只是在染色体上进行有选择的DNA复制，形成许多分叉。二氢叶酸还原酶基因扩增为第三种类型，扩增的基因以染色体外因子成对存在。

真核细胞的细胞核基因由三种RNA聚合酶转录。RNA聚合酶Ⅰ只转录rRNA操纵元，RNA聚合酶Ⅱ转录所有蛋白质基因，聚合酶Ⅲ主要转录低分子量的RNA包括5SrRNA和tRNA。这三种酶结构相似，都有2个大亚基和多个小亚基，而且都有一些共同的亚基。

三种RNA聚合酶都不能直接识别其启动子。各种转录因子识别启动子中相关序列并与RNA聚合酶结合，将RNA聚合酶准确地定位到转录起点上。各种有关转录因子通过一系列反应，在各种启动子上组装成转录启始复合体。转录因子TBP是三种RNA聚合酶启始转录时所必需的，TBP以其一个亚基结合到转录起点附近。

真核生物基因组中大多数启动子都由RNA聚合酶Ⅱ识别，转录起点上游的TATA框是这类启动子的共同结构特征。TATA框的一般作用是确定转录起点，使转录从其下游约30bp的位置开始。大多数RNA聚合酶Ⅱ的启动子还在TATA框上游有另一些重要序列如CCAAT框和GC框。

同RNA聚合酶Ⅱ的启动子一样，RNA聚合酶Ⅰ的启动子也位于基因的5'端。它含有两种对rRNA基因有效转录具有重要作用的结构，其中之一是转录起点附近的核心序列，另一结构是上游控制要素UCE。

RNA聚合酶Ⅲ的启动子有两种，典型的启动子位于转录单位以内，另一些基因启动子则位于转录起点的5'端。

增强子能极大地提高启动子的转录效率，与启动子的主要差别在于它可以从远距离作用于其上游和下游的基因。沉默子的作用方式与增强子相似，但它

抑制基因转录。RNA 聚合酶Ⅱ行使功能至少需要 6 种普通转录因子，TFⅡA、B、D、E、F 和其他基因特异性转录因子。TFⅡD 含有 TATA 框结合蛋白(TBP)。TBP 与启动子的 TATA 框结合，作为形成前启始复合体的组织中心。大多数基因特异性转录因子与某些基因的增强子结合。

RNA 聚合酶Ⅰ的前启始复合体除了含有聚合酶Ⅰ外，还含有另外两种转录因子。上游结合因子与 UCE 结合，促进 RNA 聚合酶Ⅰ结合到转录起点附近的核心要素上。另一种是含有 TBP 的 SL1 转录因子，它促进 RNA 聚合酶Ⅰ结合到启动子上。

RNA 聚合酶Ⅲ的典型的启动子依赖于两类普通转录因子，一种为组装因子，如 5S rRNA 基因转录时所需的 TFⅢA 和 TFⅢC，转录 tRNA 基因只需 TFⅢC。另一种为启始因子 TFⅢB，它只对 RNA 聚合酶Ⅲ的结合起定位作用。

促进 RNA 聚合酶Ⅱ转录的基因特异性转录因子可以与启动子成员如 GC 框或 CCAAT 框结合，也可与有关基因的增强子结合。大多数基因特异性转录因子都含有三个不同的域，即一个 DNA 结合域，一个转录活化域和一个二聚体化域。

细胞核中具转录活性的基因都将其启动子附近的序列附着到核骨架上，而不表达的基因则不与核骨架相连。核骨架的作用是为基因转录提供活化蛋白。核小体也是影响基因转录的重要因素，当转录开始以后，转录机器无须消除组蛋白 8 聚体，但组蛋白 8 聚体与启动子结合则抑制转录。

真核生物基因组中大多数蛋白质基因和 tRNA 基因以及少数 rRNA 基因都是隔裂基因，其编码序列由不编码的内含子隔开。RNA 聚合酶Ⅱ转录的初级产物为前体 mRNA，含有内含子及其他额外序列，这种初级转录子叫做不均一核 RNA (hnRNA)。前体 mRNA 须经过成熟如除掉内含子、加帽、加尾等过程才成为具翻译功能的 mRNA。

前体 mRNA 含有加工的信号，如内含子 5’端都有 GU，3’端有 AG，5’和 3’剪接位点还包括重要的、与内含子剪除有关的共有序列。一组叫做核酶的核内小分子 RNA (snRNA) 与前体 mRNA 的剪接有关，它们通过两个转酯基步骤将内含子除掉和将外显子连接在一起。

细胞核中 mRNA 的多聚 A 尾巴在转录后由多聚 A 化酶加上，mRNA 含有加尾信号序列 AAUAAA。多聚 A 尾巴防止 mRNA 降解和促进翻译。成熟 mRNA的 5’端的帽子结构由加帽酶在转录后加上。帽子结构的作用也是防止 mRNA 降解和促进其与核糖体结合，形成翻译启始复合体。

真核细胞中的 mRNA 没有 SD 序列，但是其帽子结构对 mRNA 与核糖体

结合具有重要作用。翻译启始因子 eIF4F 首先与帽子结合，然后 40S 小亚基与启始 tRNA（met－RNA）一起结合进来，同时其他翻译启始因子也结合进来，最后翻译启始复合体沿 mRNA 扫描，直到找到起点密码 AUG 为止。翻译起点上游和下游的某些碱基也影响起点密码的选择，通常－3 位为 A 或 G，＋4 位为 G。当 60S 的核糖体大亚基结合进来后，翻译便进入延长期。

真核细胞的翻译延长过程同原核细胞相似，也是通过 A 位/P 位模型延长肽链。翻译终止过程也同原核生物相似，由翻译释放因子 RF 识别终止密码并终止翻译。

真核生物的许多基因都有 2 个启动子，一个为组成型，另一个则为诱导型。所以生物体可以采用多启动子控制基因的可变转录，使同一基因在不同组织或不同类型的细胞中转录出不同 mRNA，从而合成功能相似的不同蛋白质。而且，同一基因编码的同一种转录子在不同组织中通过不同加工途径即可变剪接可以产生编码功能不同的 mRNA，从而产生功能不同，甚至功能完全相反的蛋白质。

染色质的重要组成成分组蛋白只在细胞周期的 S 期合成，细胞周期在转录和转录后水平上控制组蛋白基因表达。

mRNA 的寿命也决定自身基因的转录。半衰期长的 mRNA 的基因转录水平较高，半衰期短的 mRNA，其基因转录水平较低。mRNA 半衰期的长短与其多聚 A 尾巴的长短和内部脱稳定序列有关。

某些基因转录可以进行自动调节。滑爪蟾卵母细胞特异性的 5SrRNA 受卵母细胞特异性转录因子 TFⅢA 调节，当 TFⅢA 多于 5SrRNA 时，它与 5SrRNA 基因结合，促进转录；当 TFⅢA 全部与 5SrRNA 结合、形成 5SRNP 后，卵母细胞特异性表达的 5SrRNA 基因便停止转录。

基因在翻译水平上的控制表达是生物体发育的一种重要调控机制。血红蛋白合成时，一方面通过血红素调节其他 4 条珠蛋白链的合成，另一方面两种珠蛋白也通过翻译启始过程，协调二者的平衡合成。卵母细胞中贮存的 mRNA 决定胚胎的早期发育，但其翻译发生在受精之后，即发育阶段又调节 mRNA 的翻译。高等真核生物的不同组织和细胞中存在各种适合自身 mRNA 翻译的 tRNA，其总体含量不适合其它组织中 mRNA 的翻译。不同组织中还存在不同的翻译控制 RNA（tcRNA），它们或者与自身组织中的 mRNA 结合，促进其翻译，或者与某些 mRNA 结合抑制其翻译。

蛋白质在翻译后的变化也是重要的控制基因表达的机制。某些蛋白质在翻译后只有通过修饰才具有活性，另一些蛋白质则通过翻译后的修饰而失活，并以失活状态调节基因表达。还有一些编码酶复合体的基因是在翻译后在不同组

织中降解不同亚基，在酶的降解水平上体现基因的控制表达。这种有控制的蛋白质降解过程被叫做遍在蛋白的蛋白质调节。许多蛋白质在合成以后，必须到达细胞内外适当位置才能行使功能。还有一些蛋白质复合体是在翻译后由有关蛋白质组装成的功能单位，组装过程受侣伴蛋白的调节。

高等真核生物中许多基因的表达都受发育阶段的调节，不同组织、不同细胞表达不同基因，而且同一基因在不同组织中还能采取不同表达策略。哺乳动物的脱辅基酯蛋白基因在肝细胞中编码出一种大蛋白质，而在肠细胞中则可通过 RNA 编辑，编码出一种较小的蛋白质。脊核动物的珠蛋白基因只在红血细胞中表达，在个体的不同发育阶段，珠蛋白基因家族在表达时须从两个基因簇的一个成员转变到另一成员。

高等真核生物都有性别分化，而不同性别的生物体都能采用某些特殊机制控制某些基因的性别特异表达。果蝇的性别由性染色体（X 染色体）和常染色体共同决定，即性染色体与常染色体的比例决定性别分化，但性别的分化受一关键基因即转换基因控制。如果没有转换基因，无论 X∶A 比例如何，都只产生雄蝇。转换基因转录后只在 XX 细胞中加工出雌蝇特异性转录子，这种转录子的性别特异性加工过程受另一叫做性决定致死基因的调节，它通过两种方式控制转换基因转录子的可变剪接。

高等真核生物基因组中有一小组基因依据其亲本来源，在胚胎或个体发育期间进行单等位基因表达，表现出基因组刻印现象。在这组基因中，当一个基因通过一种性别的配子传递时，其表达活性不受影响，但是当这一基因通过另一种性别的配子传递时，其活性受到抑制而不能表达。基因组刻印受多种机制调节。母本/父本基因组之间或两个等位基因之间互作，导致一个等位基因发生差别基因表达是其机制之一。在配子形成期间，DNA 发生特异性甲基化是造成基因组刻印现象的另一种机制。此外，基因组刻印也可能受基因控制。

外部环境信号也能调节基因表达。半乳糖可促进酵母的 4 个半乳糖代谢基因 gal1、gal2、gal7 和 gal10 的表达，而葡萄糖则可抑制这些基因的表达。类固醇激素通过与哺乳动物细胞中激素受体结合后再结合到依赖于激素的增强子上，提高基因的转录水平，同时也可有选择地阻止某些 RNA 降解和改变其加工途径。植物生长素也同样能活化某些植物基因的转录，并抑制另一些基因转录。较高温度可以活化热激基因表达，刺激动物、植物和微生物合成热激蛋白。各种真核生物的热激基因在结构和功能上都有很大相似性，其启动子区中都存在若干热激控制信号序列，而且有关热激基因的编码序列都有一定程度的序列同源性。热激效应也表现在翻译水平上抑制正常蛋白质合成。

光照能诱导植物的许多基因表达，这些受光诱导表达的基因在其转录起点

上游都有一至数段与接收光信号有关的、叫做光调节子的 DNA 序列。植物细胞中有关的光受体蛋白与光调节子互作，调节基因表达。缺氧条件也能调节植物某些基因表达，它直接影响基因的转录过程，使植物合成一组缺氧多肽。

复　习　题

选择 1～7 题中正确答案：

1. 真核生物的复制子具有以下特征，它们是：
   A. 比原核生物复制子短得多，因为其中存在多个终止序列。
   B. 比原核生物复制子长得多，因为其基因组比较大。
   C. 各个复制子都是双向复制并能融合。
   D. 所有复制子都同时复制以保证整个染色体在 S 期完成复制过程。
   E. 并不是都在同时复制，在一定时间内只有小部分具活性。
2. 差异剪接产生的转录子在
   A. mRNA 的 5’非翻译区不同。
   B. mRNA 的编码区不同。
   C. mRNA 的 3’非翻译区不同。
   D. 以上所有。
3. 以下转录因子中哪些含有 TBP?
   A. TFⅡB　　B. TFⅢA　　C. SL1
   D. TFⅡD　　E. TFⅢB　　F. UBF
4. 锌指蛋白与 DNA 结合：
   A. 结合到主沟中
   B. 通过锌指的 C－基末端部分与 DNA 结合
   C. 利用其 α－螺旋区与 DNA 结合
   D. 在每一指头上通过形成两段序列特异性 DNA 结合部分与 DNA 结合
   E. 利用锌指的保守氨基酸与 DNA 结合。
5. 下列有关剪接位点的陈述中哪些是正确的?
   A. 剪接位点含有数段长的同感序列。
   B. 5’剪接位点和 3’剪接位点能够互补。
   C. 成熟 mRNA 中含有剪接位点。
6. 分支点核苷酸
   A. 几乎都是 A
   B. 随机地位于内含子中

C. 位于一段同感序列之中

D. 它通过与 3’剪按位点结合，启始第一步剪接反应。

7. 可变剪接

A. 产生由同一基因编码的同种型蛋白质变异体。

B. 使用不同的 5’和 3’剪接位点。

C. 用于在不同组织和在不同发育阶段合成不同蛋白质。

D. 它可能利用不同外显子或保留内含子。

8. 真核细胞的染色体复制时，核小体结构必须复制，请说明细胞周期中各种组蛋白的合成情况，以及复制叉中新核小体的组装模型。

9. 某二倍体生物 DNA 有 4.5×10^8 碱基对，这种 DNA 在 3 分钟内完成复制。假定复制叉每分钟向前移动 10^4bp，这种生物的基因组有多少复制子？

10. 真核生物的三种 RNA 聚合酶分别合成哪些 RNA?

11. 比较真核细胞和原核细胞的 mRNA 在结构方面的异同。

12. 真核细胞的许多 mRNA 都含有内含子，mRNA 在成熟过程中，这些内含子是如何被除掉的？

13. 比较真核细胞和原核细胞的蛋白质合成启始、合成延长和合成终止的异同。

14. 在某种此住性哺乳动物的所有体细胞中，一个 X 染色体为常染色质，另一 X 染色体则为异染色质，问哪一条染色体载有活性基因？

15. 假定以下是真核生物的由 RNA 聚合酶Ⅱ转录的某个基因的非模板链的 5’端序列：

5’AATTGCGTATAAATTCGAGACTACTTACGGCTACCCGTTAGCT　，

请确定最可能的转录起点（帽子位点）。

16. 请说明一个基因以三种不同方式编码几种不同蛋白质。

17. 激素是如何参与调节真核生物的基因表达过程的？

18. 在哺乳动物中，分别来自父、母双方的基因可以进行差异表达，造成这种现象的原因主要有哪些？

19. 举例说明可变剪接是如何控制果蝇性别分化的。

20. 指出以下序列中内含子 3’剪接位点的位置。

5’ACGUACUAACAUUCUAUUCCUUAAGUUCAUAAGUUGAGUC－3’

第五章

线粒体和叶绿体基因组的组织及表达

线粒体（mitochondrion）和叶绿体（chloroplast）是真核细胞的两种主要的进行能量转换的细胞器。线粒体存在于动、植物细胞之中，其主要功能之一是在细胞呼吸期间氧化碳水化合物，释放能量。而叶绿体则为植物细胞所特有，它通过光合作用将二氧化碳还原成碳水化合物。这两种细胞器都有其相对独立的遗传系统，由其决定的某些性状的发育叫做非孟德尔遗传（non－Mendelian inheritance）或核外遗传（extranuclear inheritance），或细胞质遗传（cytoplasmic inheritance）。但是，线粒体和叶绿体都只是通过各自的生化反应机器合成几种与自身功能有关的蛋白质，而其他一些具有细胞器特异性的蛋白质的合成则需要细胞核基因和细胞器基因进行合作。因此真核细胞在控制信息大分子的合成过程中，各个遗传系统是相互依赖的，它们共同决定真核生物各种性状的发育。

第一节　线粒体基因组

一、线粒体 DNA 的复杂性

早在 20 世纪 60 年代初期就已发现在线粒体中存在 DNA，直到目前为止，所有已知的各种生物的线粒体基因组都基本上只编码几种与呼吸作用有关的多肽和几种组成线粒体内膜的磷酸化复合体的多肽以及线粒体释翻译系统的几种成分。虽然各种生物的线粒体 DNA 都存在广泛相似性，但不同物种的线粒体 DNA 在构型、大小和碱基组成方面都变化很大（表 5－1）。多细胞动物的线粒体 DNA 很小，并都呈环状。高等植物的线粒体 DNA 是不均一的，有线状分子也有环状分子。真菌线粒体基因组大多呈环状，分子量变动于 12.8Md 至 71Md 之间。线粒体 DNA 不均一性最大的，可能是数原生生物，如莱因衣藻（*Chlamydomonas reinhardii*）线粒体基因组为环状分子，只长 5μm；草履虫（*Paramecium aurelia*）的线粒体为线状基因组，长 14μm；而锥虫（*Try-*

panosome）的线粒体基因组则由大环和小环分子组成。

真核细胞的线粒体DNA除了在大小和构型方面存在很大差别外，其碱基组成变化也很大。例如，啤酒酵母线粒体DNA的G+C含量约为20%，相当的浮力密度为1.683，而在鸭类（*Anas platyrhynchos*）中，线粒体DNA的G+C含量约为60%，浮力密度为1.711。

表5-1　线粒体DNA的大小及构型

分　类	物　　种	形状	长度（μm）	分子量（$\times 10^{-6}$）
原生动物	变形虫（*Acanthamoeba castellanii*）	环状	12.7	26
	草履虫（*Paramecium aurelia*）	线状	13.8	30～35
	四膜虫（*Tetrahymena pyriformis*）	线状	～15	～30
	锥虫（*Trypanosoma mega*）	环状	9.0（大） 0.74（小）	16.1（大环） 1.49（小）
绒泡菌	多头绒泡菌（*Physarum polycephalum*）	环状	19.1	41
真菌酵母	异酒香酵母（*Brettanomyces anomalus*）	环状	18.2	～38
	卡斯酒香酵母（*Brettamonuyces custersii*）	环状	～34	～71
	近平滑假丝酵母（*Candida parapsilosis*）	环状	11.1	23.1
	姆拉克氏汉逊氏酵母（*Hansenula mrakii*）	环状		36.4
	温奇氏汉逊氏酵母（*Hamsenula wingei*）	环状	8.2	17.3
	非洲克勒克氏酵母（*Kloeckera africana*）	环状	8.3	17.5
	乳克鲁维氏酵母（*Kluyveromyces lactis*）	环状	11.4	24.0
	啤酒酵母（*Saccharomyces cerevisiae*）	环状	21～25	46～52
	糖酵母（*Saccharomyopsis lipolytica*）	环状	～14	～29
	粟酒裂殖酵母（*Schizosaccharomyces pombe*）	环状	6.0	12.5
	光滑球拟酵母（*Torulopsis glabrata*）	环状	6.0	12.8

（续）

分　类	物　　种	形状	长度（μm）	分子量（$\times 10^{-6}$）
丝状真菌	构巢曲霉（*Aspergillus nidulans*）环状	10～10	20～21	
	红色面包霉（*Neurospora crassa*）	环状	18～22	41
多细胞动物线虫类	人蛔虫（*Ascarislumbricoides*）	环状	4.8	～10.1
昆虫类	普通果蝇（*Drosophila melanogaster*）	环状	6.2	12～35
鱼类	黑鲫（*Carrassius carrassius*）	环状	5.4	～11.3
两栖动物	滑爪蟾（*Xenopus laevis*）	环状	5.8	12.2
爬行动物	饰纹箱龟（*Terrapene ornata*）	环状	5.3	11.1
鸟类	鸡（*Gallus domesticus*）	环状	5.4	11.3
哺乳动物	人（*Homo sapiens*）	环状	～5.2	～10.9
藻类	莱因衣藻（*Chlamydomonas reinhardii*）	环状	4.5	9.8
有胚植物	玉米（*Zea mayz*）	环状		～300
	马铃薯（*Solanum tuberosum*）	线状		～100
	月见草（*Oehotera berteriana*）	环状	～31	～66
	甜瓜（*Cucumis melo*）			～1 600

二、哺乳动物线粒体基因组

根据对人体、鼠和小牛的线粒体DNA序列的比较分析，人体和小牛线粒体基因组的组成相似，二者的核苷酸序列存在70%的同源性。在人体线粒体基因组中（图5－1）已知的基因包括12S和16SrRNA基因、22种tRNA基因、细胞色素脱辅基蛋白（cytochrome B apoprotein）基因、细胞色素C氧化酶（cytochrome Coxidase）的亚基Ⅰ、Ⅱ和Ⅲ基因以及ATP酶的亚基6基因和5个NADH脱氢酶基因。植物线粒体基因组还含有5SrRNA基因，而酵母和

其他真核生物则不含 5S rRNA 基因。哺乳动物线粒体基因组还含有其他一些编码蛋白质的基因。

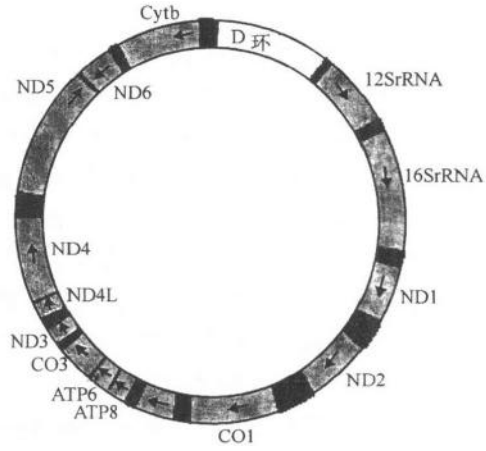


图 5－1　人体线粒体基因组的基因分布图

涂黑区示 tRNA 基因，阴影区示编码区，箭头示转录方向，CO 为细胞色素氧化酶基因，ND 为 NADH 脱氢酶基因

哺乳动物线粒体基因组非常紧凑，基因之间几乎不存在非编码序列。基因组的两条 DNA 单链在碱基组成方面不同，其中一条链的双环碱（嘌呤碱）较多，浮力密度较大，叫做重链（heavy strand，H）。另一条链嘌呤碱较少，浮力密度较小，叫做轻链（light strand，L）。大多数基因都位于 H 链上，只有几个基因（主要为 tRNA 基因）位于 L 链上。

哺乳动物线粒体基因组的重要特征之一就是 rRNA 和蛋白质基因的编码序列都由一个或少数几个 tRNA 基因隔开。这种现象表明其转录子需要进行加工。另一重要特征是许多可读框都缺少终止密码，只是 T 或 TA 位于最后一个有义密码之后或紧接在一个 tRNA 基因或另一可读框之前。

哺乳动物线粒体基因组的转录是对称和完整的，可能其基因组只有一个启动子位点，它可能位于“D 环”区。D 环区无编码功能，含有 H 链的复制启始区（见本节后）。

三、酵母线粒体基因组

在酵母中，以芽殖酵母线粒体基因组的研究最为深入，其基因组为环状分子，长 25μm（50Md、75kbp），富含 AT 碱基，其 G+C 含量很低，只有 18%。

在富含AT碱基的区域中、ATT结构发生缺失突变的频率很高，通常认为这种序列不能编码蛋白质，所以芽殖酵母线粒体基因组可能含有多个DNA复制起点。

同哺乳动物和大多数真核生物线粒体基因组一样，芽殖酵母的线粒体DNA也只编码几种重要的线粒体成分（图5－2），主要包括24种tRNA基因、核糖体小亚基（37－40S）的15SrRNA和大亚基（50－60S）的21SrRNA基因、一种与线粒体核糖体结合的蛋白质（var1）的基因、细胞色素c氧化酶三个亚基的基因（复合体Ⅳ的亚基（COⅠ、Ⅱ、Ⅲ）、ATP合成酶复合体的两个亚基基因（线粒体ATP酶的亚基6和9）以及细胞色素b的脱辅基蛋白（b－c1复合体的部分，复合体Ⅲ）基因（cob）。ATP酶亚基9又称DCCD结合蛋白（decyclohexylcarbodimide），分子量为76510D，在哺乳动物和红色面包霉中，该亚基由核基因编码。在酵母线粒体基因组中，除了一个tRNA基因以外，其他所有基因都位于同一线粒体DNA单链上。酵母线粒体基因组也有几个编码产物尚不清楚的可读框。

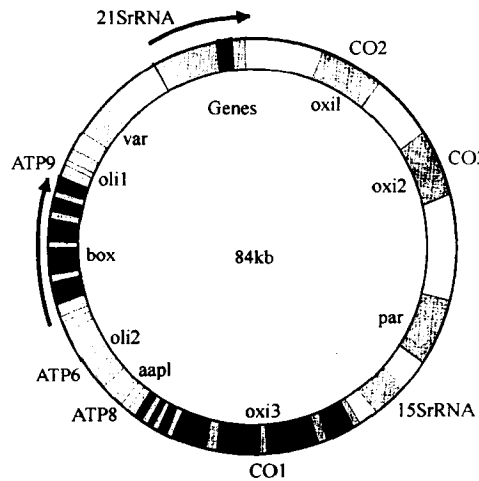


图5－2　酵母线粒体基因组的基因分布图

图中同时表示编码蛋白质的隔裂基因和非隔裂基因，涂黑区示内合子，阴影区示外显子，oli和aap示对寡霉素敏感的ATP酶亚基基因，box为cob基因，par为未知功能的可读框，var为核糖体小亚基蛋白

酵母线粒体基因组中，编码线粒体内膜多肽（脱辅基细胞色素bf、coⅠ、coⅡ、coⅢ、ATP酶的亚基6和9）的基因的位置与哺乳动物线粒体中相应基因的位置不同，而且酵母线粒体的两种rRNA基因的相对距离很远，转录方向相同；在哺乳动物线粒体中，虽然这两个基因的转录方向相反，但二者连在一

起。另一与哺乳动物不同的是，酵母线粒体 tRNA 基因并不分散在 rRNA 基因和蛋白质基因之中。

酵母线粒体基因组中有些基因是隔裂基因，其可读框含有一个以上内含子，但目前在哺乳动物线粒体基因组中尚未发现隔裂基因。在啤酒酵母的某些品系中，同一线粒体基因的内含子的有无对基因表达不具有明显影响，这种内含子是不重要的，所以叫做选择内含子（optional intron）。在 21SrRNA 基因中有一个内含子与重组极性有关。其中至少有 2 个内含子为选择内含子。coI 基因有 9 个内含子，其中有些内含子中存在可读框，其编码产物都可能参与转录子的加工过程。目前已知线粒体中至少有 9 种与 RNA 剪接有关的蛋白质即成熟酶（maturase）或反作用子由内含子编码。

四、植物线粒体基因组

各种高等植物的线粒体 DNA 都有相近的浮力密度（1.705～1.706），相当于 46% 的 G+C 含量。植物线粒体的遗传组成也相当复杂，其 DNA 分子的构型和大小因分析方法不同和植物种类不同差别很大，其中有环状 DNA，也有线状 DNA。在不同植物中，环状 DNA 分子的长度变动于 0.5～30μm 之间，碱基数相当于 200kbp～2400kbp（表 5－2）。

有些植物的线粒体基因组除了具有一个大环状的主染色体以外，还具有若干小环状和线状 DNA，其大小一般变动于 1～100kbp 之间（表 5－3）。这类小分子 DNA 叫做质粒或附加体。在玉米中，线粒体 DNA 既可以单个大环分子（1200kbp）的形式存在，也可以三个小环（600、350 和 250kbp）的形式存在。大环分子中存在 2 段长 3kbp 的重复序列，重复序列碱基配对后，就会发生分子内重组，形成 3 个小环状 DNA 分子。

表 5－2 植物线粒体 DNA 的特征

植物名称	g/cm^3	G+C（%）	基因组大小（kb）
甜瓜	1.705 0	45.4	2 400
黄瓜	1.704 5	44.9	1 500
西瓜	1.705 7	46.1	330
豌豆	1.705 5	45.9	360
玉米	1.704 6	45.0	570
南瓜	1.704 8	45.2	840

表 5－3　高等植物线粒体中小分子量 DNA

植　物	长度 (kb)	结　构	植　物	长度 (kb)	结　构
玉米			*Zea diploperennis*	7.5	线状（D－1）
N.cms－C.T.S	1.94	小环		5.4	线状（D－2）
cms－C.	1.42	小环	高粱		
	1.57	小环	cms	5.7	线状（N－1）
cms－S	0.35	线状（s_3）		5.3	线状（N－2）
	5.4	线状（s_2）	烟草（细胞培养）	10.1	寡聚体
	6.2	线状（s_1）		28.8	寡聚体
	54.0	环状	菜豆（细胞培养）	1.9	寡聚体
	117.0	环状	月见草	6.3	环状
N 型	67	环状（48%）		7.0	环状
	49	环状（20%）		8.2	环状
	99	环状（14%）		9.9	环状
	13	环状（7.6%）		13.5	环状
	23	环状（5.8%）	野油菜	11.3	线状
	136	环状（3.8%）	大豆	20	环状
	30	环状（1%）		33	环状
cms－T	82	环状（47%）		42	环状
	1.3	环状		54	环状
	1.4	环状		68	环状
	1.45	环状		81	环状
	1.5	环状		99	环状
	7.3	环状	蚕豆	1.4	环状
	56	环状（37%）		1.7	环状
Ru 型	7.5	线状（R－1）	（cms350）	1.5	环状
	5.7	线状（R－2）		1.6	环状
玉米（细胞培养）	1.5	寡聚体			
	1.8	寡聚体			

资料引自 Sederoff 等（1985）。

由于植物线粒体 DNA 的大小和结构差异很大，一般认为植物线粒体基因组比动物和酵母线粒体基因组具有较多基因或调节序列。如果植物线粒体中 3 种 rRNA 和 24 种 tRNA 各有一个顺反子，以及考虑到线粒体能够合成 20 种不同多肽，那么线粒体 DNA 中编码序列的分子量至少可高达 16×10^6Md，因此植物线粒体 DNA 含有大量基因。

在许多植物的线粒体中，各种大小不同的线粒体 DNA 之间都存在一定程度的序列同源性。如烟草线粒体中 28.8kb 的 DNA 分子的所有限制性片段都与 10.1kb 的线粒体 DNA 存在同源性。又如，玉米线粒体中 1.94kb 的质粒不仅与其 S_1 和 S_2 以及线粒体主染色体存在同源性，而且还与细胞核 DNA 存在同源性。从进化的观点分析，植物线粒体中各种 DNA 分子可能起源于一个单环分子。在进化过程中，单环分子通过一系列重复、缺失、重组等遗传变异，从而产生各种大小不同的、含有重复基因和非编码序列的 DNA 分子。

某些植物线粒体 DNA 还与其细胞核 DNA 存在同源性。在玉米线粒体 DNA 中，与细胞核 DNA 存在同源性的序列主要分布在一些小质粒如 S_1 和 1.94kb 的质粒等之中。

高等植物线粒体基因组中还普遍存在叶绿体 DNA 序列。在玉米基因组中主要有三类叶绿体 DNA（图 5－3）。第一类为 12kb 的颠倒重复序列，其中包括丙氨酸 tRNA 、异亮氨酸 tRNA、缬氨酸 tRNA 和亮氨酸 tRNA 的编码序列以及 16S rRNA 的编码序列等。第二类是长度为 1.9kb 到 2.7kb 的序列，其中

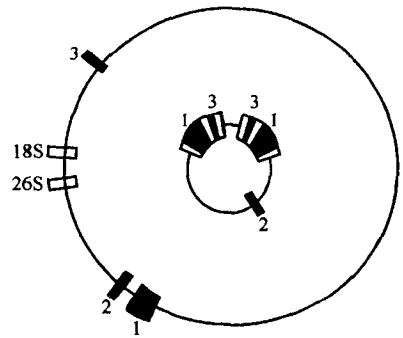


图 5－3　玉米线粒体主染色体（外环，570kb）和叶绿体染色体（内环，139kb）中的同源序列

（1）为 12kb 的叶绿体颠倒重复序列，（2）Rubisco 大亚基基因，（3）为 23S、4.5S 和 5S rRNA 基因

（根据 Lonsdale 1985）

包函 Rubisco 酶大亚基的整个编码序列。第三类是长度为 2kb 的序列，在叶绿体基因组中，这段序列编码 23S rRNA 的一部分，4.5S 和 5S rRNA 以及精氨酸 tRNA，天冬酰氨 2 和甘氨酸 1 的 tRNA。此外，玉米线粒体 S1 DNA 还含有叶绿体 psbA 基因的一部分序列。

由于线粒体和叶绿体基因组的遗传密码不同，转录启始和终止信号不同以

及翻译信号不同，所以一般认为线粒体中的叶绿体基因不具有功能。但是，由于线粒体和叶绿体 DNA 的碱基组成差别很大（叶绿体的 G+C 含量为 37%，线粒体为 46%），可能叶绿体 DNA 对线粒体 DNA 的稳定性以及复制和转录过程具有某种位置效应。例如玉米线粒体基因组中 12kbp 的叶绿体颠倒重复 DNA 部分缺失后，会造成细胞质雄性不育（cytoplasmic male sterelity，CMS）。

总的说来，线粒体基因组中与叶绿体和细胞核 DNA 存在同源性的序列可能是一类调节序列，在植物的发育过程中可能与各种细胞器以及细胞核之间通讯联络有关。

五、原生动物线粒体基因组

在低等真核生物中有一类原生动物叫做锥虫（*Trypanomatidae*），有时也称为动基体原生动物（kinetoplastid protozoa）。这类低等真核细胞只有一个线粒体，其线粒体 DNA 叫做动基体 DNA（kinetoplast DNA，kDNA）。锥虫线粒体 DNA 以拟核体（nucloid body）形式存在于鞭毛基部线粒体的基质中。kDNA 由 20～50 个大环分子和 5000～10000 个小环分子组成。这许许多多的环状分子连接在一起，形成一种网络。大环分子的大小为 20～39kb，小环分子为 0.5～2.5kb。在每一网络中，大环 DNA 序列均相同，小环 DNA 序列则是不均一的。

锥虫的线粒体基因主要位于大环 DNA 分子内，包括编码 2 种线粒体 rRNA 的基因、呼吸链中的几种成分如细胞色素 b（cytoshrome b，cyb），细胞色素氧化酶亚基Ⅰ、Ⅱ、Ⅲ（coⅠ、coⅡ、coⅢ），以及 NADH 脱氢酶亚基 1、4、5（NDI、4、5）。此外，还有 4 个尚未确定的可读框。大环还含有大量可变区，其长度因种或变种不同而有所不同。可变区大多由富含 AT 碱基的重复序列组成，这类序列可能不编码蛋白质，而编码与 mRNA 编辑有关的指导 RNA（guide RNA，gRNA）。目前尚未发现锥虫线粒体基因组中存在 tRNA 基因，线粒体中的 tRNA 可能都由核基因编码。

gRNA 是锥虫线粒体中一类小分子 RNA，长约 60 个核苷酸，能够与线粒体 mRNA 进行碱基互补配对。gRNA 除了能与 mRNA 进行正常的 A－U 和 G－C 碱基配对以外，还能进行非正规的 G－U 和 C－A 碱基配对。

目前知道小环分子编码 gRNA，其大小、组织及序列同源性在同一物种中都有差别，但在不同种之间，小环分子的变异与 RNA 编辑有关。

六、线粒体基因密码的不通用性

所有生物的线粒体蛋白质基因都有不使用通用密码的现象，也就是说，即使线粒体基因与核基因和原核生物基因的三联密码的碱基相同，但由线粒体基因的同一密码编码的氨基酸与核基因和原核生物基因密码编码的氨基酸不同（表 5－4）。例如，哺乳动物和酵母线粒体基因都采用终止密码 UGA 编码色氨酸，采用编码异亮氨酸的正常密码 AUA 编码甲硫氨酸，正常情况下甲硫氨酸由 AUG 编码。另一方面，不同生物的线粒体基因中，编码同一种氨基酸的密码也不相同，如动物和酵母线粒体基因用终止密码编码色氨酸，而植物线粒体基因则可以用正常密码 UGG 编码色氨酸。最后，在某些低等真核生物如锥虫以及高等植物线粒体中，线粒体 DNA 中的密码与成熟 mRNA 分子中的相应位置的密码是不相同的。例如在植物线粒体基因中，CGG 密码在由这些基因编码的氨基酸序列中相应位置的氨基酸或者是色氨酸，或者是精氨酸，所以过去曾认为线粒基因的 CGG 密码可以编码色氨酸。但后来发现，成熟 mRNA 中编码色氨酸的密码都是正常密码 UGG，即 DNA 分子中 CGG 密码在 mRNA 中由 UGG 所取代。造成这种密码转换的机制目前尚不清楚，可能线粒体通过一种特殊的 RNA 编辑机制将碱基 C 转换成 U。

表 5－4　线粒体基因与核基因密码功能的差异

密　码	核基因	酵　母	哺乳动物
CUU	亮氨酸	苏氨酸	亮氨酸
CUC	亮氨酸	苏氨酸	亮氨酸
CUA	亮氨酸	苏氨酸	亮氨酸
CUG	亮氨酸	苏氨酸	亮氨酸
CGG	精氨酸	色氨酸	色氨酸
UGA	终止密码	色氨酸	色氨酸
AUU	异亮氨酸	亮氨酸	亮氨酸
AUA	异亮氨酸	甲硫氨酸	甲硫氨酸
AUG	甲硫氨酸	中间甲硫氨酸	甲硫氨酸
AGA	精氨酸	精氨酸	终止密码
AGG	精氨酸	精氨酸	终止密码

七、线粒体基因组的复制

有关各种生物的线粒体基因组的复制机制目前都还不清楚。根据对哺乳动物线粒体的研究，线粒体中存在一种线粒体特异性的 DNA 聚合酶即 γ－DNA 聚合酶。线粒体 DNA 在复制时，γ－DNA 聚合酶首先复制其中一条链即 H 链（图 5－4）。当 H 链复制叉前进到基因组 1/2 处时，再开始以 L 链为模板的复制过程。线粒体中也存在拓扑异构酶，这种酶能够催化切割和复制终止过程，即将两条亲链解螺旋和最后将两条子双螺旋分开。

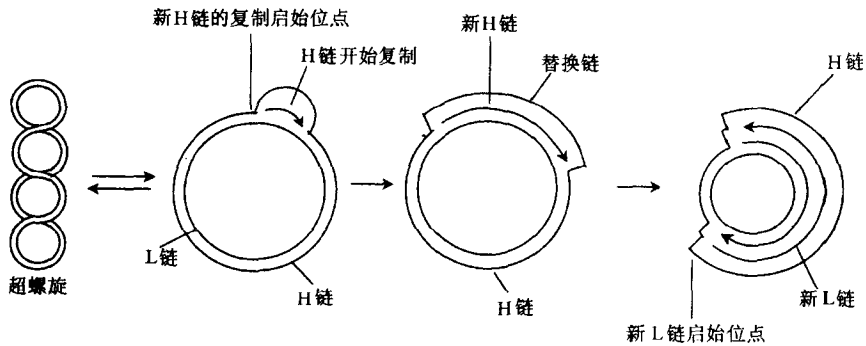


图 5－4　线粒体 DNA 的复制模型

首先复制 H 链，当替换环扩大到环状线粒体 DNA 的 1/2 处时，L 链开始复制

第二节　线粒体基因的转录与翻译

一、线粒体基因的转录和转录控制

（一）转录酶及其识别信号

有关线粒体基因转录机制的研究主要来自酵母。酵母线粒体只含有一种 RNA 聚合酶，由核基因编码，由一个 145kD 的核心亚基和至少一个转录特异因子即启动子识别因子组成。酵母线粒体 RNA 聚合酶的核心亚基不具有转录特异性，并不与启动子进行特异结合。这种核心亚基与噬菌体 T_3 和 T_7 的 RNA 聚合酶存在很高的同源性，不同的是线粒体聚合酶在识别启动子时需要一种特殊因子即一种 43kD 的蛋白质。这种特殊亚基自身无转录活性，也不与 DNA 结合，只是起着促进核心亚基启始转录的作用。现在还发现酵母线粒体中

存在另一种转录特异因子即 70kD 的 DNA 结合蛋白。

酵母线粒体基因组有 19～20 个转录启始位点，其同感序列为 ATATAAGTA，其中 3'端的一个核苷酸相当于 mRNA5'端的起点核苷酸。这个 9 核苷酸序列存在于所有转录起点上游，但对启动子功能具有关键作用的序列为 -10～+2 之间的序列。

（二）转录调节

酵母线粒体基因的转录各不相同，部分原因可能是不同基因的启动子强度不同，其强度有 20 倍之差。启动子强度不仅受核苷酸序列的影响，而且也受其相对位置的影响，即一个强启动子的顺式作用能抑制另一较弱启动子的活性。影响线粒体基因转录的另一重要因素是启动子在多顺反子转录单位中的位置。当然线粒体基因转录也明显受环境条件的影响。

二、线粒体转录子的加工

（一）某些内含子编码与转录子加工有关的剪接蛋白

如前所述，虽然哺乳动物线粒体基因不具有内含子，但许多低等真核细胞的线粒体基因以及高等植物的线粒体基因都有内含子，包括 rRNA 基因和一些编码蛋白质的基因。根据对酵母和植物线粒体突变体的研究，线粒体转录子的加工过程通过几种途径进行，其剪接因子包括由核基因编码的蛋白质、内含子中特殊序列以及由内含子本身或其他基因的内含子编码的多肽。

由内含子编码的蛋白质参与转录子的加工过程是线粒体转录子剪接的一个独特方面。在酵母线粒体的某些隔裂基因中，从第一个外显子开始的可读框从剪接位点开始，一直延伸到内含子序列。例如酵母的细胞色素 b 脱辅基蛋白基因 cob 一般由 6 个外显子和 5 个内含子组成，但内含子数目因酵母品系不同而有不同。其中有些内含子含有可读框，编码几种加工酶也叫成熟酶，以促进自身转录子的内含子以及其他转录子的内含子的切割（图 5-5）。cob 转录子在加工过程中首先切割内含子 1，将开头的两个外显子连接在一起，然后 RNA 进行翻译，形成由 2 个 5' 外显子和内含子 2 同共编码的融合蛋白。融合蛋白长 351 个氨基酸，其中由外显子 E_1 和 E_2 编码的部分长 143 个氨基酸，由内含子可读框编码的部分长 208 个氨基酸。融合蛋白然后再参与内含子 3 的切割。由于这种多肽破坏自身 mRNA，所以这种现象称为剪接自动调节或剪接稳恒（splicing homeostasis）。cob 基因的内含子 4 也有一个可读框，也编码成熟酶，这种多肽促进该内含子自身及 co Ⅱ基因的内含子的切割。在酵母线粒体基因组中，除了 cob 基因的内含子有可读框外，其他基因的内含子也有可读框，如

coI 基因以及 21SrRNA 基因的内含子都具有编码某种多肽的可读框，其中 coI 基因的内含子 1 与 cob 基因的内含子 4 存在 50% 的同源性，可能其编码产物也参与前体转录子的剪接过程，因为在这个内含子中发生突变后，酵母细胞就不能合成具有功能的细胞色素氧化酶亚基 I。

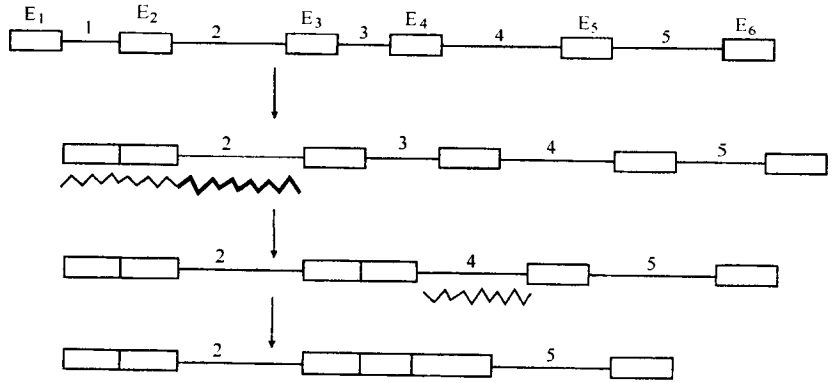


图 5－5　酵母 cob 基因前体转录子中内含子的剪除步骤

E 及方块示外显子，线条及阿拉伯数示内含子，曲线示多肽

（二）内含子的自动剪接

根据对细菌和各种高等真核生物的细胞核 RNA、线粒体以及叶绿体 RNA 的内含子序列的比较分析，各种基因的内含子可以分成三组，即在第四章中所介绍的细胞核前体 mRNA 的内含子、见于细胞器 RNA 和细菌 RNA Ⅰ组内含子的 G1（group 1 intron）和Ⅱ组内含子 G2（group 2 intron）（G1 内含子也见于某些低等真核生物的细胞核基因）。G1 和 G2 内含子在序列同源性、次级结构以及加工机制方面都有明显差别，但其切割过程都不需要蛋白质因子或核酶的催化作用（图 5－6）。

1. Ⅰ组内含子的切除　Ⅰ组内含子的内部序列可以通过互作，将内含子折叠成绕同一核心的次级结构，这类内含子普遍见于多头绒泡菌的细胞核 rRNA 基因、酵母线粒体的 21SrRNA 基因和 coI 基因以及玉米和菜豆叶绿体 tRNA 基因等。同细胞核蛋白质基因的内含子剪接过程一样，Ⅰ组内含子的剪接也是通过转酯基作用即磷酸酯基的转移完成的，但不同的是线粒体内含子的剪除不需各种 UsnRNA。在离体条件下，进行Ⅰ组内含子的自动切割时，需要加入亚精胺（spermidine），以提高 G1 内含子切割和抑制其它自动切割。G1 内含子自动切割过程需要鸟嘌呤核苷三磷酸，因为游离的 GTP 的 3'羟基启始转酯基反应的第一步（图 5－6），它与内含子 5'切割位点形成氢键，切割 5'外显子。

根据对酵母21SrRNA剪接过程的研究，切割下来的内含子在5'端具有碱基G，但这一碱基并不由DNA编码，而是在切割过程中加到内含子5'端的，因

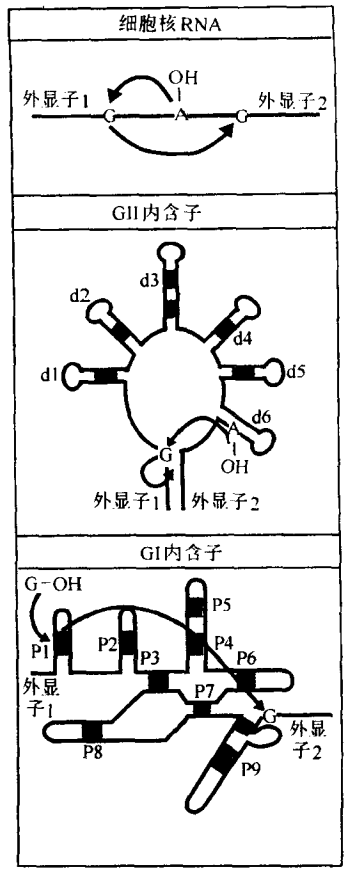


图5-6　三种内含子都通过2个连续的转酯基过程被除掉

第一步反应中，一个游离的羟基连接到5'外显子/内含子相邻处，在细胞核和G2的含子中由RNA分子内部的2'-OH提供，在G1内含子中则由一个游离的鸟嘌呤核苷酸提供。在第二步反应中，释放下来的外显子的3'-OH连接到3'内含子/外显子连接处，P为配对的区域，D为GⅡ内含子的亚区

为基因组中内含子5'端无碱基G，只有U。

G1内含子能够直接催化自身的切割过程，由于转酯基作用首先需要进行磷酸基转移，如果没有游离的GTP或其他核苷三磷酸，前体mRNA链上任何2'羟基或3'羟基都可以提供这种转移功能。如果使用2'羟基，剪接反应可以产生套环结构。

并非所有 G1 内含子在离体条件下都能自动剪除，可能许多内含子在切割过程中需要某些特殊蛋白质，否则就不能折叠成适合自动切除的正确结构。前述某些线粒体 mRNA 中，由内含子编码的“成熟酶”可能帮助 RNA 折叠成具有活性的构型。

2.Ⅱ组内含子的切除　线粒体Ⅱ组内含子具有与Ⅰ组内含子不同的保守序列和次级结构，与细胞核 mRNA 内含子的结构比较相似。在离体条件下，酵母线粒体的切割下来的内含子呈稳定的环状结构，其自动切割需要 Mg^{++}，但不需要 GTP 或其他任何核苷三磷酸。当然，自动切除反应同样需要亚精胺以提高 G2 内含子切割和抑制其他切割。自动切割反应还需要严格的温度值(45℃)。切割下来的内含子或呈环状，或呈线状。环状结构是通过与分支点结合而形成的，即内含子的 5’磷酸基与分支点碱基 A 形成 2’－5’磷酸二酯键。线状分子可能是被打破的内含子套环。

线粒体 mRNA 的 G2 内含子的切割和外显子连接与细胞核前体 mRNA 的剪接机制基本一致（图 5－6)。酵母线粒体 coI 基因的内含子同细胞核 mRNA 内含子一样，内含子套环的分支点 A－G 连接也是通过 2’－5’磷酸二酯键形成的，但不同的是线粒体 G2 内含子的切割不需要 UsnRNP，某些线粒体 G2 内含子切割需要“成熟酶”。

G1 和 G2 内含子的切割机制基本相似，都包括 2 个步骤。第一步切割 5’剪接位点，第二步切割 3’剪接位点。这两个切割反应都产生 5’磷酸和 3’羟基末端，而且两个切割步骤都与连接有关，所以反应中磷酸二酯键的数目无变化。但是，G1 内含子切割的第一步涉及鸟嘌呤的 3’羟基引起的磷酸基转移，而 G2 内含子切割或细胞核 mRNA 剪接的第一步则涉及 RNA 链内部核苷酸 2’羟基引起的磷酸基转移。第二步在两个系统中都非常相似。

三、RNA 编辑

RNA 编辑（RNA editting）由 R.Benne（1986）根据对锥虫线粒体基因的研究提出，最初是指锥虫线粒体 RNA 以一种特殊的加工方式，将尿嘧啶核苷插入 RNA 中或从 RNA 中除掉，最后将不具有翻译功能的 RNA 变成可翻译的 mRNA。后来发现其他低等真核细胞的线粒体基因、高等植物的线粒体基因以及某些病毒基因在表达过程中，其成熟 mRNA 的碱基序列也并不能反映相应 DNA 的碱基序列。所以后来就将 RNA 编辑定义为细胞中某些特殊的翻译控制装置将不具功能的转录子加工成可翻译的 mRNA，或者是将已具功能的 mRNA 进行修改，使 DNA 分子中不完整的遗传信息在 RNA 水平上保持完整的一种

特殊加工机制。在不同生物的线粒体中，RNA 编辑可能涉及不同机制，而且编辑形式包括碱基插入、缺失、转换和颠换（表 5－5）。例如在锥虫线粒体的细胞色素氧化酶亚基 coⅡ和Ⅲ基因、细胞色素 b（cyb）基因以及 NADH 脱氢酶的三个亚基基因中的 ND1 基因等都在可读框的不同位点发生碱基 U 的插入或缺失。哺乳动物的某些基因则表现为 C→U 转换，多头绒泡菌线粒体则表现为 C、A 或 U 插入，小麦线粒体的细胞色素 C 氧化酶（coxⅡ）基因以及 ND_3 基因等则表现为 C→ U、G→A 和 U→A、G→U 的颠换。但目前尚未发现在酵母线粒体中存在 RNA 编辑现象。

表 5－5　RNA 编辑的不同类型

编辑类型	机　　制	物　　种
U 的插入和缺失	gRNA（转酯基作用?）	锥虫线粒体
C、A 或 U 插入	未知	粘菌线粒体
G 插入	不清	附粘病毒 P 基因
3’ 端加 A	多聚 A 化	脊椎动物线粒体
C→U 转换	脱氨基作用	哺乳动物脱酯基蛋白 B 基因
C→U 转换或相反	脱氨基/氨基化作用?	高等植物线粒体

注：根据 Cattaneo，R，1991。

在锥虫线粒体中，RNA 编辑与 gRNA 有关。gRNA 能够与待编辑的 mRNA 进行碱基配对，形成杂二聚体分子。杂二聚体分子中，除了正常的 A－U、G－C 碱基对以外，还包括非正常的碱基对即 G－U 和 C－A 碱基对。有关 gRNA 在 RNA 编辑过程中的作用目前有两种比较流行的假说。第一种为误配对识别假说（mismatch recognitoin），这种假说认为编辑机器在 gRNA 和 mRNA 二聚体的 5’端即从 mRNA 的 5’端多轮次地切割误配对的碱基对，然后再将碱基 U 插入或将 U 从 mRNA 分子中除掉，以消除错误配对的碱基，直至 gRNA 完全与 mRNA 进行碱基配对。第二种假说为所谓的配对保护模型（match protection model），这种模型认为，在未编辑的 RNA 的某些特定区域，尿嘧啶的缺失和插入是随机的，mRNA 中已编辑的部分通过与 gRNA 形成二聚体，避免进一步被编辑。

关于编辑机器的各种部件，目前除了知道其中的 gRNA 和 mRNA 以外，尚未发现参与这一过程的其他分子。可能还有其他一些大分子复合体参与 RNA 编辑过程，这类复合体可能同 RNA 的内含子剪接复合体一样，也由 RNA 和蛋白质组成。在 gRNA－mRNA 二聚体中，人工合成的 gRNA 的 3’多聚 U 尾巴能够与 mRNA 的内部区域相连，因此可能碱基 U 的插入和缺失通过转酯基作用完成（图 5－7）。gRNA 可能识别 mRNA 及其中被编辑位点和指导碱基 U 的插入或缺失，而复合体则可能含有催化分子、RNA 移位分子或稳定 gRNA－mRNA 二联体的分子。另一方面，由于在两种锥虫（*T. brucei* 和 *Leishmania tarentolae*）的线粒体抽提物中都测定到连接酶活性，可能 gRNA－mRNA 杂种分子不是通过转酯基作用形成的，而是通过内切核酸酶切割和连接酶连接形成的（图 5－8）。当然，锥虫线粒体中究竟是通过转酯基作用还是通过

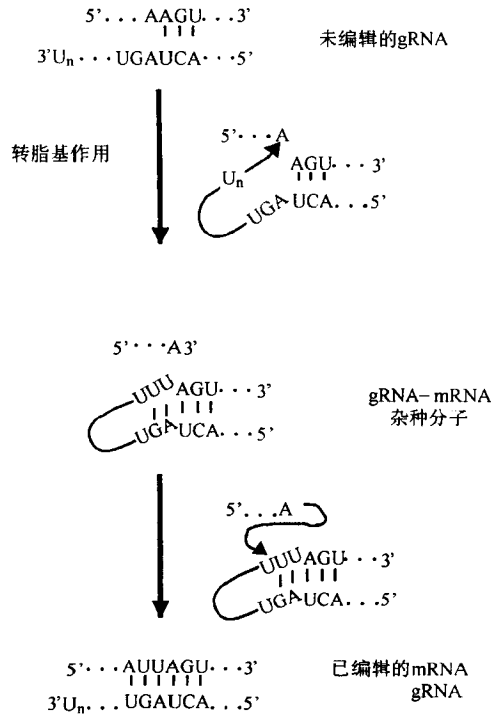


图 5－7　gRNA－mRNA 杂种分子的形成及其切割连接反应

gRNA3’多聚 U 尾巴首先与 mRNA 中编辑位点共价结合（向上箭头），形成中间杂种分子，然后释放的 mRNA 片段的 3’端可能直接连接到二聚体的 gRNA 的寡聚 U 中，释放 gRNA 和将部分编辑的 mRNA 连接在一起（向下箭头）

切割连接反应形成 gRNA－mRNA 杂二聚体，有待进一步研究。

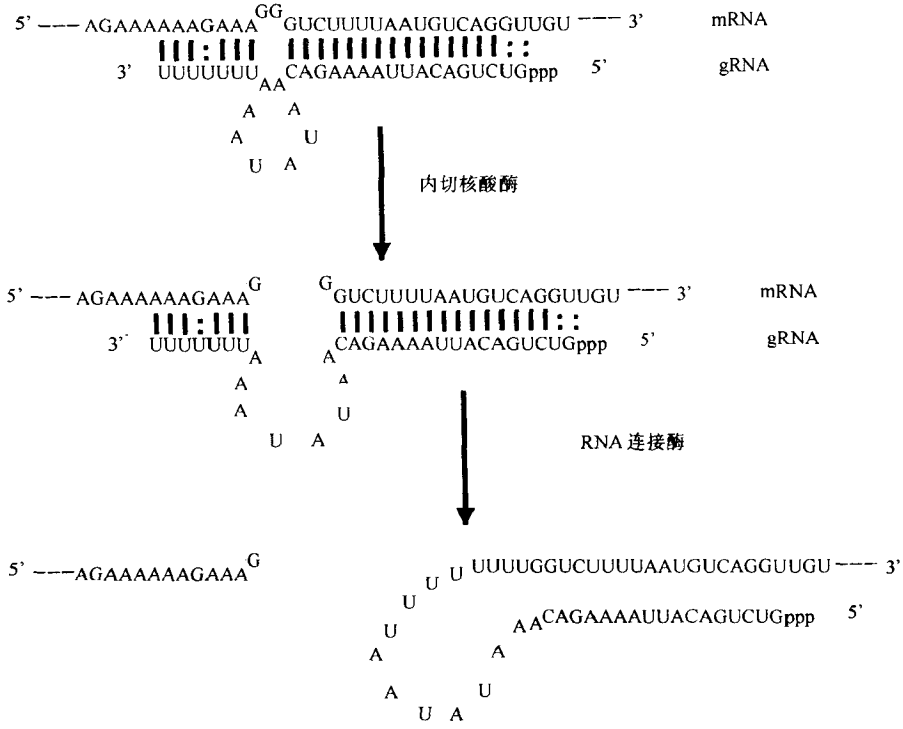


图 5－8　形成 gRNA－mRNA 杂种分子的切割连接机制

上图为离体合成的 CYbmRNA 和 CYb－gRNA，实线为正常碱基对，冒号为 GU 碱基对，下图为可能形成的 gRNA－mRNA 杂种分子

（根据 Simpson，L．等 1995）

四、线粒体 mRNA 的翻译和翻译控制

（一）核基因编码与线粒体 mRNA 翻译有关的蛋白质

根据对酵母线粒体基因表达的研究，线粒体 mRNA 的翻译受许多细胞核基因的控制。这些核基因编码线粒体翻译系统的一般成分包括线粒体核糖体蛋白以及 tRNA 合成酶等。此外，线粒体 mRNA 的翻译还至少需要三种由核基因编码的正活化因子。

酵母线粒体核糖体由 77 种不同多肽组成，除核糖体小亚基蛋白 Var1 蛋白外，其他所有亚基都由核基因编码。目前还不清楚在其他生物的线粒体中，是

否所有核糖体蛋白也都由核基因编码或由核基因和线粒体基因共同编码。

在酵母中，细胞核基因还编码 10 种线粒体氨酰基 tRNA 合成酶，其中有 8 种只在线粒体中行使功能，另外 2 种与线粒体 tRNA 和细胞溶质 tRNA 合成有关。

酵母线粒体中 2 种类似大肠杆菌翻译延长因子 Tu 和 G 的多肽也由核基因编码，这 2 种翻译调节因子只在线粒体中行使功能。此外酵母核基因组中也有 2 个编码线粒体翻译抑制子的基因。

（二）tRNA 与线粒体遗传密码

各种生物的线粒体 tRNA 都比细胞溶质 tRNA 的分子量小，而且数目也少。线粒体 tRNA 的 3'－CCA 序列不由基因组编码，是在转录后通过加工过程接到 tRNA 上的，这一特征与细胞溶质 tRNA 相同。

线粒体 tRNA 中许多都是同功 tRNA，所以线粒体只用少数几种 tRNA 就能将大量遗传信息翻译出来。大多数生物的线粒体中只有 24 种 tRNA，这 24 种 tRNA 能够将线粒体上 32 种密码翻译成蛋白质。如果利用线粒体外翻译系统，至少需要 32 种 tRNA，所以一种 tRNA 可以阅读一个以上密码。在哺乳动物线粒体中，24 种 tRNA 的反密码子可以与 60 种密码配对，因此每一种 tRNA 可以阅读 2～4 种同义密码。例如反密码子 UAG 可以与 CUU、CUC、CUA 和 CUG 配对，反密码子的第一个碱基 U 可以与 U、C、A 和 G 配对。

（三）翻译启始复合体形成及特异调节

关于线粒体翻译启始的一般特征目前知道的不多，不同生物的线粒体中形成开放翻译启始复合体时核糖体的沉降系数各不相同。如酵母线粒体核糖体为 74S（大亚基 50～60S，小亚基 37～40S），高等植物线粒体核糖体为 77～78S（大亚基 50S，小亚基 33S）。在酵母线粒体中，大多数 mRNA 的 5'端都有一段长数百个碱基的前导序列，其中有一至数个起点密码 AUG，但无固定的核糖体结合位点。3'端也接有一段非翻译序列，其中含有一段保守的 12 核苷酸序列 AAUAAUAAUCUU。在形成翻译启始复合体时，尚不清楚核糖体究竟选择哪一个 AUG 作为起点密码。酵母线粒体的另一基因 cox Ⅱ 的 mRNA 非常特殊，其前导序列很短，而且无 AUG 起点密码。

对于高等动、植物线粒体而言，一般认为线粒体翻译启始复合体的形成类似细胞溶质翻译系统，即首先形成小亚基/mRNA/tRNA 复合体，然后大亚基再结合进来，启始翻译过程。

在细胞溶质 mRNA 的翻译过程中，启始复合体的形成与 mRNA5'端帽子结构以及 3'端多聚 A 尾巴有关。根据对 Hella 细胞线粒体 mRNA 的分析，其 5'端没有帽子结构，只是 3' 端有长约 55 个碱基的多聚 A 尾巴，但这种多

聚 A 尾巴并不由基因组编码，而是在转录后加到 mRNA 3'端的。但是具有翻译功能的植物线粒体 mRNA 既无帽子结构，也无多聚 A 尾巴。可能帽子结构和多聚 A 尾巴对翻译启始复合体形成并不具有重要作用。

另外，根据对酵母线粒体翻译系统的研究，线粒体中存在由核基因编码的、具有 mRNA 特异性的翻译活化子。这些活化子能够特异性地促进 cox3mRNA 的翻译，其中已知的有 PET494、PET54 和 TET122 三个翻译活化子。另一个基因 cob 的 mRNA 的翻译也依赖于由核基因编码的活化子（BSI）和（BS2）。由核基因编码的翻译调节子可能通过与 mRNA 的 5'和 3'非翻译区结合，稳定 mRNA 和调节 mRNA 的可翻译状态。

第三节　植物质体基因组的组织及表达

所有植物的分生细胞都有一些小的、未分化的、叫做前质体（proplastid）的细胞器，随着细胞分化，前质体也就分化成各种特征不同的质体，如淀粉体（amyloplast）、叶绿体（chloroplast）、有色体（chromoplast）等。虽然各种不同类型的质体在大小、超微结构、色素形成、生化活动以及功能方面各不相同，但它们都具有相同的生化反应装置，即质体遗传系统。叶绿体是植物细胞中最常见的一种质体，它除了具有进行光合作用的整个生化反应装置以外，还参加氨基酸、核酸、脂肪以及淀粉的合成，所以对叶绿体的遗传研究最为深入。

一、质体基因组的组织和结构

根据对叶绿体的分析，质体基因组为双链环状 DNA 分子，通常以多拷贝形式存在。其拷贝数因植物组织或细胞不同差别很大。在幼叶中，每一个叶绿体或白色体就含有 200 个拷贝以上的质体染色体。在叶片中，质体 DNA 的含量约占 DNA 总含量的 10%～20%，而在根组织中，大约只占 1%左右。质体染色体的大小也因物种不同变化很大（表 5－6），某些绿藻（*Codium*）只有 50kbp，某些巨型绿藻的叶绿体基因组可以大到 2000kbp，但高等植物的叶绿体染色体一般变动于 120～160kbp 之间。虽然高等植物的叶绿体 DNA 都是均一的，但某些褐藻（*Pylaiella littoralis*）的叶绿体基因组则由两种不同的环状 DNA 分子组成，一种为 133kb，另一种为 58kb。

大多数植物的叶绿体 DNA 在氯化铯中的浮力密度为 1.679g/ml，相当于 37%的 G+C 含量。虽然 G+C 含量因物种不同有一定变化，但通常只变动于 36%～40%之间。高等植物的质体 DNA 不含 5－甲基胞嘧啶，也不与组蛋白结

表 5－6　质体染色体的大小和构型

分　类　单　位	染色体大 IH（kb）	构形	重复序列大小（kb）
被子植物			
230 个种 90 个属 30 个科	120～217	单环	10～76
裸子植物			
银杏（*Ginkgo biloba*）	158	单环	17
蕨类植物			
挂皮紫萁（*Qsmonda cinnamonea*）	144	单环	10
雀巢蕨（*Asplenium nidus*）	150	单环	–
凤尾蕨（*Pteris vitata*）	150	单环	–
苔藓植物			
地钱（*Marchantia polymorpha*）	121	单环	11
圆顶褒果苔（*Sphaerocar piusdomellii*）	125	单环	–
绿藻			
松藻（*Codium* sp）	57	单环	–
刺松藻（*Codium fragile*）	85	单环	–
椭圆小球藻（*Chlorella ellipsoidea*）	174	单环	23
莱因衣藻（*Chlamydomonas reinhardii*）	195	单环	22
史密氏衣藻（*C. smithii*）	199	单环	23
卵配衣藻（*Chlamydomonas eugametos*）	243	单环	20
摩氏衣藻（*C. moewusii*）	292	单环	41
素衣藻（*Polytoma obtusum*）	200	–	–
地中海伞藻（*Acetabularia mediteranae*）	2 000	–	10
裂隙伞藻（*A. cliftonia*）	2 000	–	–
其他藻类			
网地藻（*Dictyota dichotoma*）	123	单环	–
无柄无隔藻（*Vaucheria sessilis*）	125	单环	–
胞内蓝藻（*Cyanophora paradoxa*）	127	单环	10
纤细裸藻（*Euglena gracilis*）	130～152	单环	6
黑顶藻（*Sphacelaria* sp）	150	多环	–

合形成复合体。在大多数植物中，叶绿体 DNA 都有一对很长的颠倒重复，其长度在 6～76kbp 之间。颠倒重复由一大一小两个单拷贝区隔开（图 5－9）。某些植物如豌豆、蚕豆、苜蓿和松树的叶绿体 DNA 没有颠倒重复。还有一些植物的叶绿体 DNA 具有顺式重复，如纤细裸藻叶绿体 DNA 含有三段顺式重复，每一重复片段都含有一个 rRNA 基因簇。所以，叶绿体基因组可以分成三类，第一类缺少颠倒重复，第二类具有颠倒重复，第三类具有顺式重复。

一般说来，在植物的同一分类学单位种以内质体 DNA 基本相同。后来根据大量 DNA/DNA 分子杂交证明，即使在毫无亲缘关系的物种之间，质体 DNA 也至少有 30％的序列同源性。例如烟草、矮牵牛、黄瓜以及滨黎属植物

的叶绿体 DNA 几乎完全同源，玉米与菠菜叶绿体 DNA 只在 50kb 的倒位区段内存在一个很小的次级倒位区段的差别，其他区域也几乎完全相同。

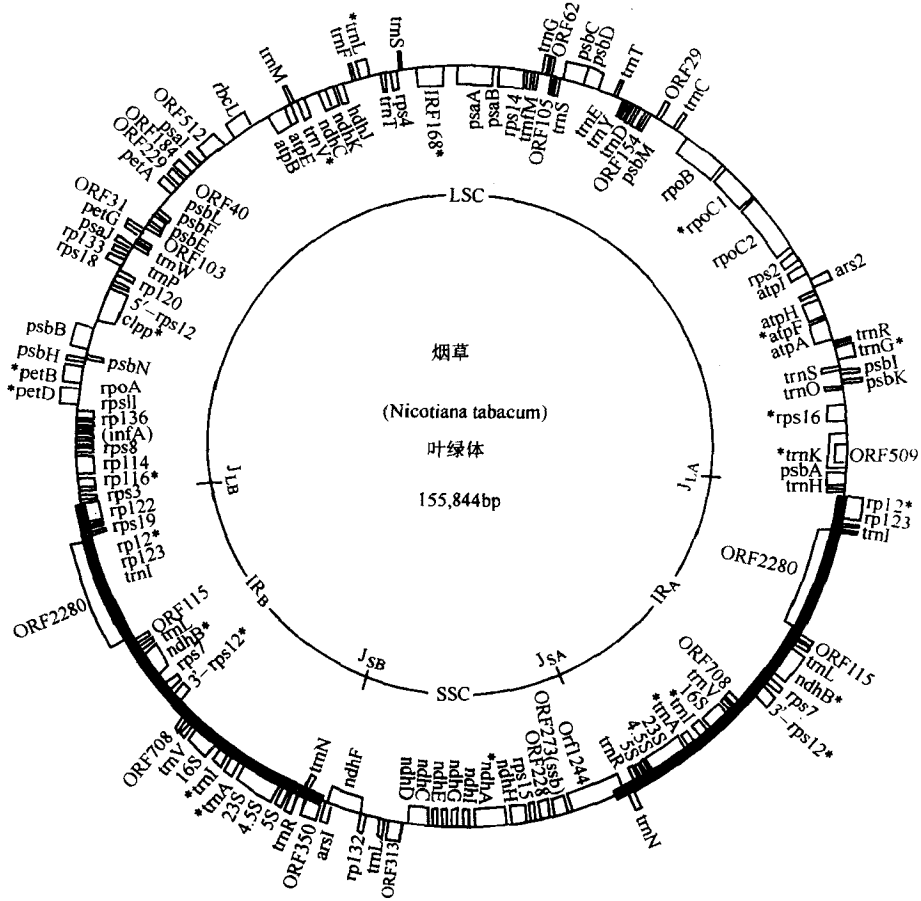


图 5－9　烟草叶绿体基因组的基因图

环内为顺时针方向转录的基因，环外为反时针方向转录的基因，星号示隔裂基因，IRF 为有内含子的可读框，IR 为颠倒重复，LSC 为大单拷贝区，SSC 为小单拷贝区，J 为 IR 和 LSC/SSC 之间的连接区

（根据 M. Sugiura，1992）

叶绿体所载有的基因基本上与线粒体相似，只是所含基因更多（表 5－7）。叶绿体基因组含有其蛋白质合成所需的所有 rRNA 和 tRNA 基因，约 50 种蛋白质基因，包括 RNA 聚合酶的部分亚基和核糖体的部分亚基的基因。所有这些基因都是由质体基因表达机器进行转录和翻译的。

表 5－7　叶绿体基因组所含基因的种类

基　　因	种　类　数
编码 RNA 的基因	
16S rRNA	1
23S rRNA	1
4.5S rRNA	
1	
5S rRNA	1
tRNA	30
类囊体膜系统蛋白的基因	
光合系统 I（psa）	5
光合系统 II（psb）	12
细胞色素 b/f（pet）	4
H^+－RTP 酶	6
基因表达系统的基因	
核糖体蛋白	21
RNA 聚合酶亚基	3
其他蛋白质基因	2
NADH 脱氢酶	6
铁氧还原蛋白	3
Rubisco 酶大亚基	1
其他未知	23
总数	120

叶绿体核糖体的沉降系数为 70S，其 50S 大亚基含有 23S、5S 和 4.5S rRNA，其中 4.5S 的 rRNA 只存在于高等植物的叶绿体中，30S 的小亚基只含有 16S rRNA。在烟草、玉米等植物的叶绿体 DNA 中，编码这 4 种 rRNA 的基因以操纵元的形式组织在一起，受一个转录开关控制，其顺序一般为 16S－23S－4.5S－5S rRNA 基因。在各种陆地植物中，tRNA 基因分散在整个叶绿体基因组中，其中有些还分布在 rRNA 操纵元的内含子和前导序列中。但在纤细裸藻中，大多数 tRNA 基因均以基因簇形式存在。有些植物的叶绿体基因组还具有 tRNA 假基因。叶绿体 tRNA3’端的－CCA 序列不由基因组编码，同

细胞溶质 tRNA 和线粒体 tRNA 相似。

在编码蛋白质的基因中，有 26 个基因与叶绿体基因表达有关，27 个基因编码类囊体膜蛋白，10 个基因与电子传递有关，还有 23 个功能尚不清楚的可读框。

二、质体基因的表达

（一）转录酶、启动子及转录终止信号

叶绿体至少含有两种不同的 RNA 聚合酶，一种为可溶性酶，另一种叫做转录活性染色体（transcriptionally active chromosome，TAC）。可溶性 RNA 聚合酶由质体基因和核基因共同编码，质体基因组中有 4 个基因即 rpoA，rpoB，rpoC1 和 rpoC2 分别编码分子量为 38kD，120kD，78kD 的 180kD 的 RNA 聚合酶亚基。TAC 则是首先从纤细裸藻叶绿体中分离到的一种疏水成分，它同叶绿体染色体紧密结合。与可溶性 RNA 聚合酶的主要差别是，TAC 不转录外源模板。RNA 聚合酶与 TAC 一起紧密地结合到叶绿体 DNA 上，倾向性地转录 rRNA 基因。

叶绿体 RNA 聚合酶必须在其他一些转录辅助因子的帮助下才能准确地启始转录过程。在玉米叶绿体中这种辅助因子叫做 S 因子（S factor），在菠菜和芥菜叶绿体中则叫做 BF 成分（BF fraction）或称类似 σ 因子的多肽。

根据在离体条件下利用大肠杆菌 RNA 聚合酶和叶绿体 RNA 聚合酶对叶绿体 DNA 转录过程的研究，知道许多基因的转录起点上游都存在类似大肠杆菌“－10”和“－35”框的 DNA 序列（图 5－10）。几乎所有叶绿体基因的“－10”框的最后一个碱基都与大肠杆菌“－10”框的最后一个碱基相同，二者都为 T。在“－35”位上，叶绿体基因也都具有高度保守的、与原核生物基因相同的 TTG 序列。但是，有些叶绿体 tRNA 基因在离体条件下转录时并不需要 5’转录控制区。在菠菜叶绿体中，atpB 基因的转录子则完全不具有非翻译前导序列，转录起点位于翻译起点密码上。菠菜叶绿体还有一个 atpB 基因缺少明确的“－10”框。在高等植物中，叶绿体 psbA 基因不仅具有原核生物基因的“－10”和“－35”框，而且在这两个框之间还存在一个类似细胞核基因 TATA 框的结构。TATA 框对 psbA 基因的准确转录也起着重要作用。总的说来，叶绿体基因组至少含有三种结构不同的非转录控制区，即含有“－35/－10”框的启动子区、无上游启动子和类似 TATA 框的启动子的非转录控制区。这些特征表明，叶绿体中可能存在多种 RNA 聚合酶和多种类似 σ 因子的辅助因子。

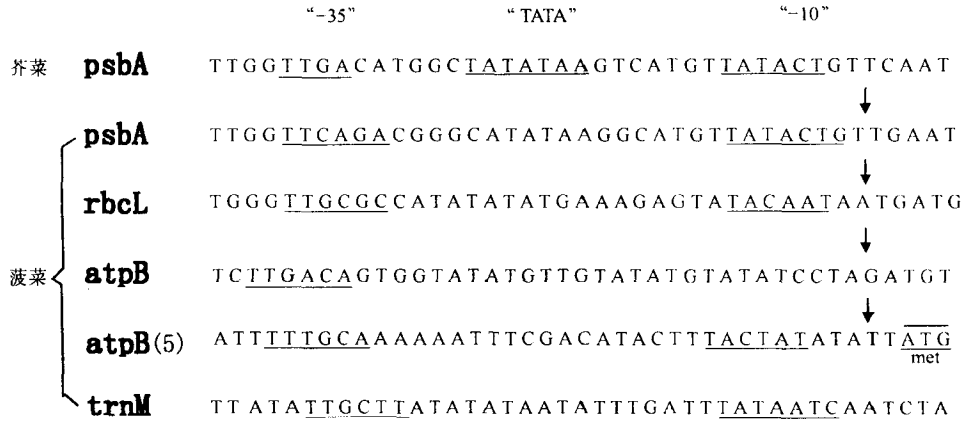


图 5-10　叶绿体基因的启动子区

其中第 5 个 atpB 基因的转录起点就在翻译起点密码之内

在原核生物基因中，转录终止位点下游都存在若干能够形成发夹结构的序列，或者是 T 或者是 A 碱基串。某些叶绿体基因也具有这种特征，但无 A 碱基串，如玉米、菠菜和豌豆的 rbcL 基因，菠菜和烟草的 psbA 基因等。玉米的 rbcL 基因在其转录终止区存在 2 个发夹环状结构，总长度为 98bp，这段序列可能是该基因转录的直接终止信号，但也有可能是 RNA 的加工信号。另外，虽然菜豆的 rbcL 基因也具有发夹结构，但实际转录终止位点却在这种结构下游数百个碱基的位置，可能这种发夹结构不能独立地终止转录过程，它还需其他某种辅助终止序列的帮助。

（二）叶绿体转录子的加工

叶绿体基因组含有 120 个以上的基因，但大约只有 50 个转录单位，因此许多基因都是连在一起进行转录，形成多顺反子 RNA。多顺反子 RNA 以及许多隔裂基因的前体转录子都必须通过加工才能形成具有功能的 RNA 分子。已知在叶绿体基因组中有 2 个操纵元由几种功能不同的基因组成，一个是 psbDC 操纵元，这一操纵元含有 ps Ⅱ 系统的两个基因和一个 tRNA 基因。另一个是 psaA 操纵元，这一操纵元含有 psI 系统的两个基因和一个核糖体蛋白基因。在各种叶绿体操纵元中，有些操纵元进行组成型转录，因此对多顺反子进行转录后的加工可能是叶绿体基因控制表达的一个重要步骤。

叶绿体多顺子 RNA 通常在转录后都须经过加工，形成许多较短的重叠 RNA 分子，其中有些较短的 RNA 为单顺反子 RNA，另一些则仍为多顺反子

RNA。例如菠菜的 psbB 操纵元含有 4 个基因即 psbB－psbH－petB－petD，其初级转录子通过切割可以形成 18 种主要 RNA，其中光合系统Ⅱ的两种 RNA（psbB 和 psbH）为单顺反子 mRNA，但编码 pet 亚基的 RNA 则为二顺反子 mRNA。这些单顺反子和二顺反子 mRNA 是两类主要的具有翻译功能的 mRNA。但是玉米叶绿体的 psbB 操纵元的转录子则都是多顺反子，可能这种多顺反子 mRNA 在翻译过程中无须将其间隔序列除掉。

大多数叶绿体转录单位在其 3'端都有若干短的颠倒重复。如前所述，这些颠倒重复可能是转录终止信号，由于在离体条件下这些颠倒重复不能终止转录过程但能够有效地、准确地启动 RNA 加工过程，所以这类序列也有可能是 RNA 的加工信号而不是转录终止信号。

现在知道，由核基因编码的某些蛋白质参与叶绿体 mRNA 的成熟过程和有选择地稳定某些叶绿体 mRNA。如由核基因编码的 28kd 的蛋白质与叶绿体 mRNA3'部分某些特殊序列结合，催化 mRNA3'末端的加工过程。

许多叶绿体基因都有内含子，包括莱因衣藻的某些 rRNA 基因、高等植物叶绿体的 tRNA 基因以及植物和藻类的某些蛋白质基因。在高等植物中，大多数具有内含子的基因都只有一个内含子，但纤细裸藻的很多蛋白质基因都有多个内含子。由于内含子的边界序列及其可能形成的次级结构不同，叶绿体基因的内含子切除可能涉及多种不同的加工机制，其中某些初级转录子的剪接可能采用与核 RNA 相同的剪接机制。

非常特殊的是，有些叶绿体基因可以分割成几个独立转录的部分，各部分转录以后，不同转录子再通过剪接连接在一起，形成具有翻译功能的 mRNA。像这样一类基因叫做分段基因（divided gene），其转录子的这种剪接方式叫做反式剪接（trans－splicing）。例如，烟草的一种编码叶绿体核糖体蛋白 CS_{12} 的基因分裂成三个部分，即一个 5'rps12 拷贝和 2 个 3'rps12 拷贝。5'rps12 含有由 38 个密码组成的外显子 1，而 3'rps12 则由 78 个密码的外显子 2 和一段 536bp 的内含子以及一个有 7 个密码子的外显子 3 组成。两个 3'rps12 分别位于叶绿体颠倒重复区内，具有相反的转录极性，5'rps12 距最近的一个 3'rps12 为 28kb(图 5－11)。这三个部分都独立地进行转录，产生两种转录子，然后这两种转录子再通过反式剪接，形成编码 CS12 蛋白的成熟 mRNA。根据烟草 rps12 基因的结构可以看出，其转录子的加工既需要进行前述顺式剪接，同时也需要进行反式剪接。

莱因衣藻的 psaA 基因也是分段基因，该基因也被分割成三部分(图 5－11)。第一个外显子有 30 个密码，第二个外显子有 60 个密码，二者相距 50kb，第三个外显子有 661 个密码，距第二个外显子 90kb。这三个外显子都分别地进

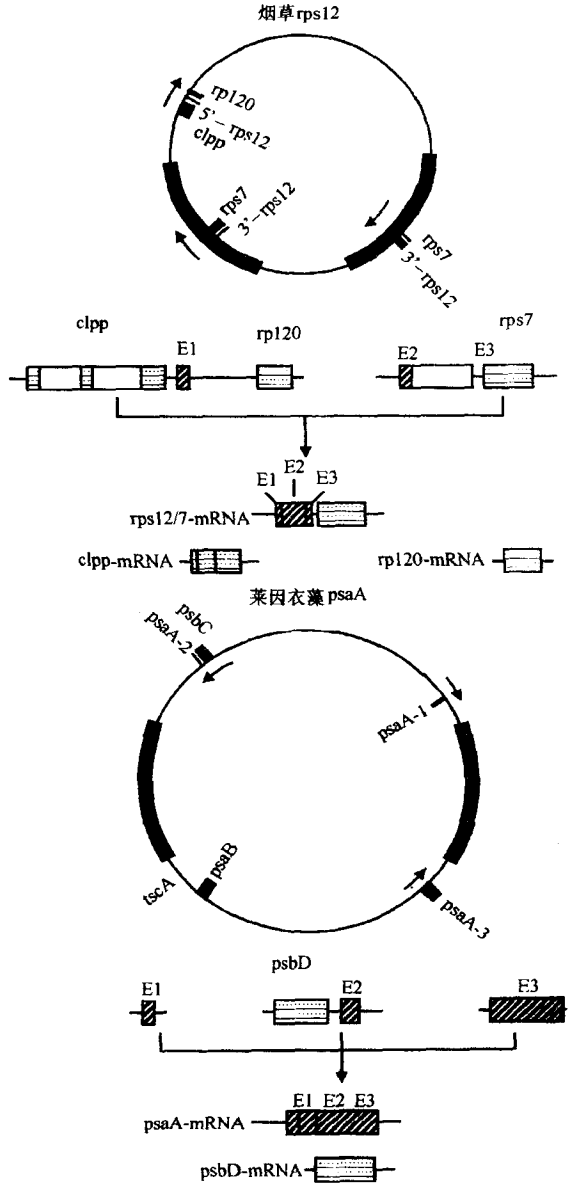


图 5－11　叶绿体分段基因的前体 mRNA 的反式剪接

环状结构示基因在染色体上的位置和转录方向以及共同转录的其他基因；下图示初级转录子及其加工后形成的成熟 mRNA，E 为 rps12 和 psaA 的外显子，未涂阴影的框为内含子

（根据 M.Sugiura，1992）

行转录，然后三条前体转录子再通过反式剪接，组装成成熟的 psaA mRNA。莱因衣藻叶绿体基因组中至少有一个基因编码对外显子 1 和 2 进行反式剪接的蛋白质，这种蛋白质使二者连接成 430 个碱基的小 RNA 片段。

(三)mRNA 的翻译启始

在叶绿体中，大多数 mRNA 的 5'区域都存在非翻译前导序列，其长度从几个核苷酸到数百个核苷酸不等，只有个别基因例外。一般认为 mRNA 的前导序列与核糖体结合有关。在原核生物中，前导序列通常形成次级结构，行使翻译调节功能。在陆地植物和衣藻中，叶绿体 mRNA 的 5'和 3'非翻译区都能形成次级结构，这两种次级结构可能都行使翻译调节功能。它们是由核基因编码的蛋白质的结合位点。

关于叶绿体翻译启始复合体的组成目前不清楚，但现在知道叶绿体 90%以上的基因(如玉米、菠菜和豌豆的 rbcL 基因)的 mRNA 都具有类似原核生物 mRNA 的核糖体结合位点。由于 rbcL 基因的 mRNA 能够在大肠杆菌的表达系统中有效地翻译成 Rubisco 的大亚基，所以一般认为前导序列中类似原核生物 mRNA 的核糖体结合位点的结构能够有效地启始翻译过程。另一方面，某些叶绿体基因如 atpB 基因的翻译起点密码上游并不存在类似原核生物 mRNA 的核糖体结合位点的序列，而且不同物种中该 mRNA 前导序列的保守性都很低，还常常发生大量插入和缺失突变，因此 mRNA 中与核糖体结合的功能可能由起点密码下游、紧邻起点密码的保守序列所决定。

第四节　线粒体和质体基因组与细胞核基因组的互作

一、细胞器的生物发生与核基因组的关系

线粒体和叶绿体并不是一个完全独立的生化反应系统，除了其基因组的表达高度地受到细胞核遗传系统的控制外，而且这两种细胞器的生物发生也是核基因与各自的细胞器基因协调互作的结果。对线粒体而言，这种细胞器在生物发生过程中，其外膜蛋白完全由核基因编码，而内膜蛋白则由核基因和线粒体基因共同编码(图 5－12)。在由线粒体基因编码的、为数有限的蛋白质中，主要为内膜系统的三种酶复合体的疏水亚基，而大多数线粒体蛋白质包括与线粒体基因组复制、转录和翻译有关的酶或蛋白质都由细胞核基因编码。

如前所述，酵母细胞色素 c 氧化酶复合体的 9 个亚基中有 3 个亚基即 Co Ⅰ、Ⅱ、Ⅲ由线粒体基因编码，其余 6 个亚基由核基因编码。ATP 合成酶复合

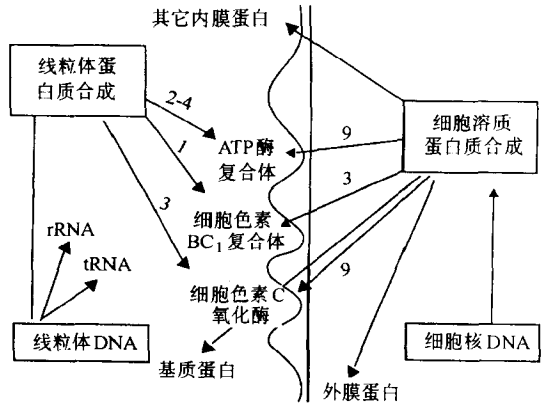


图 5－12　线粒体基因组和核基因组在线粒体的生物发生过程中的互作

数字示线粒体内膜上各种酶复合体的亚基数

体的 9 个亚基中大多数亚基都由核基因编码，只有少数亚基由线粒体基因编码。ATP 合成酶由两部分组成，即埋在膜中的 F_0 部分和暴露在膜外的 F_1 部分。在人体、小鼠和小牛线粒体基因组中只有一个编码 MF_0 部分的第 6 和第 9 亚基的基因。而在植物叶绿体类囊体膜中的 ATP 酶的 9 个亚基中，有 6 个亚基由质体基因编码，3 个亚基由核基因编码。由核基因编码的亚基为 CF_0 部分的亚基Ⅱ和 CF_1 部分的亚基 D 和 G。细胞器膜系统中另一种重要的膜蛋白为细胞色素复合体。已知在哺乳动物和果蝇线粒体的细胞色素 BC_1 复合体中，只有一个 bk 亚基由线粒体基因编码，而其它几种细胞色素 b 和细胞色素 c 成分则都由核基因编码。植物叶绿体类囊体中的细胞色素为细胞色素 BF 复合体，由 6 种成分组成，已知其中有 4 种成分由叶绿体基因编码，其他 2 种亚基可能由核基因编码。

除了线粒体和叶绿体的生物建成需要核基因组的参与以外，细胞器中其他大多数可溶性功能蛋白也都是由核基因和细胞器基因共同编码的。这些核基因产物也都是在细胞溶质的 80S 核糖体上合成，然后再运输到细胞器中组装成功能复合体，执行特殊的细胞器功能。

二、细胞溶质蛋白向细胞器的运输

细胞溶质蛋白进入细胞器的机制与蛋白质穿越内质网、进入质膜或被分泌到质膜外的机制有所不同。蛋白质穿越内质网是在与膜结合的核糖体上进行的

共翻译分泌过程，而进入细胞器的细胞溶质蛋白是在游离核糖体上翻译的，因此细胞溶质蛋白向细胞器的运输涉及到蛋白质对目标的识别和蛋白质本身的折叠问题。由于折叠的蛋白质不能通过细胞器膜系统，因此细胞溶质蛋白在翻译后，必须保持一种适合运输的构型。细胞溶质中有一类由核基因编码的抗折叠蛋白阻止目标蛋白质折叠。现在知道在真核细胞中有 2 种抗折叠蛋白，它们都属于侣伴蛋白的范畴。一种见于兔网织细胞，叫做前序列结合因子（presequense binding factor，PBF）。PBF 为一种 5.5S 的寡聚体，由数个 50kD 的亚基组成。另一种为 70kD 的热激蛋白（HSP 70），广泛存在于各种真核细胞的细胞溶质以及线粒体和叶绿体中。在酵母中，如果细胞中 HSP 70 含量很低，细胞就会累积未切割的线粒体蛋白而至死亡。

根据对哺乳动物和酵母细胞的研究，细胞溶质中的线粒体蛋白在输入到线粒体内之前，首先与线粒体外的受体分子结合。细胞溶质蛋白对细胞器的识别功能由蛋白质的前序列即运输肽的 N 基末端部分提供。线粒体外膜可能具有几种不同受体，如酵母和红色面包霉线粒体外膜上有一种分子量为 70kD 的受体，这种受体将其 N－末端疏水序列固定到外膜上，将约 60kD 的羧基末端部分朝向细胞溶质。这种受体对 MF_1 的 ATP 酶 β 亚基的输入具有促进作用。酵母线粒体的另一输入受体为 32kD 的膜蛋白，这种蛋白质与线粒体外膜上的前体蛋白形成复合体，能够促进某些前体蛋白进入线粒体。红色面包霉的 19kD 的线粒体外膜蛋白也是一种输入受体，能够促进各种前体蛋白（包括自身）进入线粒体。

过去曾认为前体蛋白质只能通过线粒体上固定通道即内、外膜相连部位进入线粒体。现在知道，线粒体内、外膜具有不同的蛋白质移位通道，都由叫做 ISP_{42}/Mom38 的蛋白质复合体和其他蛋白质亚基组成。内、外膜上的通道孔可以进行互作，它们是不固定的。线粒体基质中的侣伴蛋白通过移位通道将蛋白质拉入基质中。这种通道很大，甚至可以使与前体蛋白相连的 DNA 双螺旋通过（图 5－13）。

根据对酵母的研究，当前体蛋白的氨基末端部分进入线粒体基质后，就与线粒体中 70kD 的热激蛋白（Mhsp70）结合，并通过水解 ATP 而被释放到基质中，所以 Mhsp70 也是一种侣伴蛋白。酵母线粒体中还存在另一种侣伴蛋白 Mhsp60，以 20S 的复合体存在，由 14 个相同 60kD 的亚基组成。这种热激蛋白也由核基因编码，也是酵母细胞生存所必需的。Mhsp60 并不调节前体蛋白的实际移位过程，只是在基质中催化寡聚复合体的组装以及其他线粒体蛋白的组装。最后，许多前体蛋白的氨基末端部分都由一种叫做金属结合蛋白酶（metalloprotease）的可溶性酶进行切割，形成成熟蛋白质。

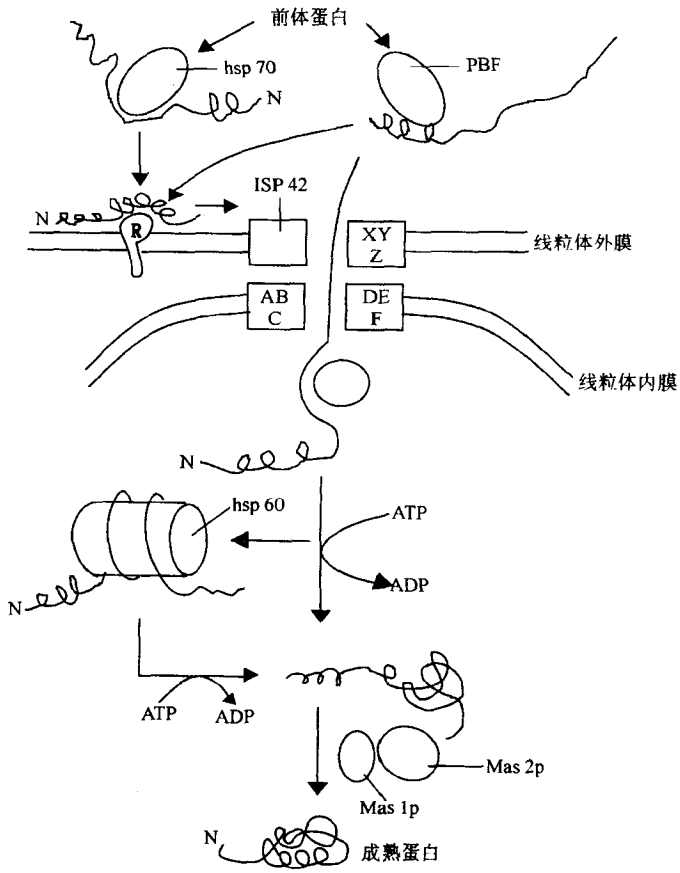


图 5－13　细胞溶质前体蛋白通过其前导肽的目标信号被运输到线粒体基质中

线粒体中由核基因编码的蛋白质很多，线粒体在输入这些蛋白质的过程中，需要对其分类和进行分配。关于这种分配机制目前有两种模型。第一种模型认为，在进入基质的蛋白质中，有些蛋白质与相应的亚基结合，组装成内膜中的复合体，另一些蛋白质则将其成熟构型转变成可溶性基质成分，还有一些蛋白质则通过输出机制运输到内膜之外，进入膜间区。进入膜间区的蛋白质分两次切割其前导肽，第一次是在进入线粒体基质时切割其前导肽的一部分，另一部分则作为输出到膜间区的信号，在内膜外表第二次切割前导肽，最后形成具有成熟 N 基末端的可溶性膜间蛋白。

第二种模型认为，细胞溶质蛋白质在通过线粒体的移位通道的过程中，有一部分蛋白质在其一部分通过外膜后，由某种转移中止信号中止其穿越内膜的移

位，前体蛋白的其余部分则继续穿过外膜，进入膜间区。与此同时，到达内膜的N基末端部分则朝向基质并进行水解加工，形成新的N基末端。这种中间产物或者被固定到移位机器上，或者在内膜外表第二次切割前导肽，形成可溶性膜间蛋白。其他完全进入线粒体基质的蛋白质的水解加工则完全与第一种模型相同。

细胞溶质蛋白向叶绿体运输的机制也基本上与向线粒体运输的机制相似（图5－14）。许多细胞溶质蛋白首先通过叶绿体的内、外膜进入基质，这一过程与蛋白质进入线粒体的过程相似。但是还有某些蛋白质需进一步穿越类囊体膜

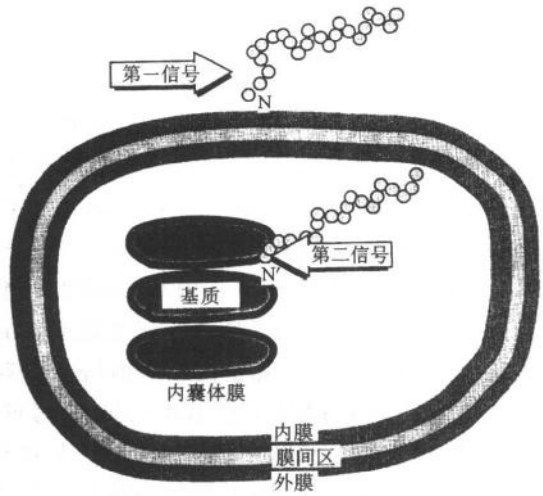


图5－14　细胞溶质前体蛋白进入叶绿体的过程

蛋白质的前导序列的N基末端部分提供穿越内、外膜或通过内、外膜进入基质的功能，在通过内、外膜期间切割前导肽的N基末端部分，前导肽的余下部分则为蛋白质通过类囊体膜提供功能

的片层进入泡腔。这些定位于类囊体膜或泡腔的细胞溶质蛋白都必须途经基质。

进入叶绿体的蛋白质的定位信号与线粒体的定位信号相似，都由叫做运输肽的前导肽提供。前导肽长约40～60个氨基酸，其氨基末端半部分用以识别叶绿体外膜。当蛋白质通过内、外膜后，便从前导肽上切割掉20～30个氨基酸。这段具有新的N基末端的前导肽用于指导对类囊体膜的识别。对于叶绿体基质蛋白如Rubisco酶的小亚基来说，当这种蛋白质通过内、外膜后，叶绿体基质中的可溶性加工蛋白酶就将前导肽全部除掉。除掉前导肽后的小亚基为一种疏水蛋白，长约110～126个氨基酸。在组装成完整的Rubisco酶之前，小亚基和

由质体基因编码的大亚基都须同某种侣伴蛋白结合，然后大、小亚基才结合在一起。

综上所述，控制细胞溶质蛋白运输到线粒体和叶绿体中的关键是前导序列，其氨基末端半部分及前导序列中另一部分序列是将蛋白质定位到外膜、膜间区和内膜中所必需的。线粒体和叶绿体与细胞核之间除了需要引入由核基因编码的蛋白质以外，这三种细胞器之间还可能需要进行核酸分子的交换，但这种交换的机制目前不清。

三、细胞器基因与核基因表达的协调性

细胞器基因组和细胞核基因组共同组成真核细胞中一个完整、协调的遗传体系，它们通过控制细胞内各种代谢活动，共同决定细胞的生长、繁殖和发育。但是，细胞器和细胞核基因组中共同编码某种蛋白质复合体亚基的基因在基因数目上存在巨大的不平衡性。例如，叶绿体中与 CO_2 同化有关的一种关键酶核酮糖 1,5 －二磷酸酯羧化酶－氧化酶复合体(Ribolose－1,5－bisphosphate carboxylase/oxygenase，Rubisco)是一种由质体基因和核基因共同编码的可溶性基质蛋白，由 8 个 55kD 的大亚基和 8 个 12kD 的小亚基组成。在高等植物和衣藻中，大亚基由叶绿体基因(rbcL)编码，小亚基由核基因(rbcS)编码。如果按每个植物细胞含有 50 个叶绿体，每个叶绿体含有 100～200 个拷贝的质体染色体计算，那么每一叶绿体基因大约重复 5000～10000 次。但是在细胞核基因组中，rbcS 基因只有几个拷贝。尽管 rbcL 和 rbcS 基因在数目上存在巨大差异，但是叶绿体中 Rubisco 酶的大小亚基是严格按比例组装的。线粒体和叶绿体中其他蛋白质复合体的合成也都与 Rubisco 酶的合成情况相似。现在一般认为，核基因表达的转录调节和细胞器基因的翻译调节是协调这些基因表达的主要机制。

现在已有大量证据表明，由细胞核基因编码的产物中除了一部分产物参与细胞器基因的转录(如细胞器 RNA 聚合酶) 和转录调节外 ，还有一部分核基因产物通过与细胞器 mRNA5'和 3'非翻译区结合控制细胞器 mRNA 的剪接、稳定和翻译。

核基因产物对细胞器 mRNA 的翻译控制大致上可分成三类。第一类是基因特异性控制(gene specific control)，即核基因编码的蛋白质与细胞器中特定基因的 mRNA 的 5'非翻译区互作或与某些细胞器 mRNA 的编码序列互作，调节细胞器 mRNA 的翻译。

第二种为类型特异性控制(class specific control)，即核基因产物协调各组对特殊环境刺激作出反应的基因的表达。例如在衣藻中，当叶绿体中总体蛋白合

成降低时，叶绿体核糖体蛋白的合成就高于各种光合作用蛋白的合成水平，但编码这两类蛋白质的 mRNA 的累积量正常。同时，与各种光合蛋白 mRNA 和核糖体蛋白 mRNA 5' 非翻译区结合的、由核基因编码的蛋白质也随之发生变化。显然，调节细胞器 mRNA 翻译的核基因的调节表达也同样表现出类型特异性。

第三种为一般控制(general control)。这种控制机制可能适用于细胞器中大多数 mRNA，其关键作用是稳定细胞器中可翻译的 mRNA，从而稳定蛋白质合成水平。一般控制的证据来自两方面，其一是核基因编码的蛋白质与线粒体和叶绿体 mRNA3' 非翻译序列结合，稳定细胞器 mRNA 和阻止早熟翻译。其二是当线粒体或叶绿体中总体 mRNA 量降低或 mRNA 过量时，可翻译 mRNA 水平保持稳定。

另一方面，环境信号也可协调细胞器中和细胞溶质中蛋白质合成。例如酵母中与线粒体生物发生有关的细胞核基因对两种主要信号即氧气和碳源作出反应。正常情况下，氧气反应由正铁血红素(heme)间接调节，而碳源葡萄糖则对基因表达有抑制作用，这种抑制作用通过一种复合信号进行传递。这些信号的传导终点为一组叫做 hap 的基因。hap 基因编码一组转录活化子，其中 HAP_1 蛋白调节正铁血红素反应，HAP2、3、4 则调节碳源反应。与此同时，编码细胞器翻译活化子的核基因也同样对这些信号作出反应。这种现象表明，环境信号在促进编码线粒体蛋白质复合体亚基的核基因表达的同时，也促进编码细胞器翻译正调节子的核基因表达，并通过翻译正调节子来协调核基因产物与线粒体基因产物的等量合成。

在线粒体和叶绿体中，由细胞器基因和许多核基因编码的细胞器成分可能都起源于共同的原核生物祖先，在后来长期的进化过程中其遗传体系才形成不同的调节表达机制。在这一过程中，一方面要保持这两套不同细胞器基因与核基因的协调表达，同时细胞器基因组中这些基因的表达也须置于核基因的控制之下。

四、植物线粒体与细胞质雄性不育性

植物细胞质雄性不育性(cytoplasmic male sterelity，CMS) 是另一说明细胞器基因组同细胞核基因组进行协调互作的例子。CMS 是一种由线粒体基因所决定的遗传现象，主要表现是雄性生殖细胞不育而雌配子可育，但细胞质雄性不育性受细胞核可育基因的抑制，如果将细胞核可育基因引入不育植株的遗传背景中，可使其后代的育性恢复正常。在玉米中有一种 S 型雄性不育系，不育植

株可以自发地恢复可育性。这种自发的育性恢复特点是其细胞核基因发生突变的结果。

根据对玉米S型不育系线粒体DNA的分析，其不育性主要由线粒体中2条线状的附加体S_1和S_2所决定(表5-3)。S_1和S_2在结构上有很多相似性，二者都具有末端重复，都与某种蛋白质共价结合。这两种S附加体可以与环状的线粒体主染色体进行重组，使主染色体变成线状分子。S_1和S_2的复制受细胞核基因控制，当雄性不育系恢复可育性时，可育株就不能产生S_1和S_2附加体。

除了玉米的S型不育系的雄性不育性状与线粒体附加体或质粒有关以外，在玉米的其他类型不育系以及高粱、向日葵、甜菜、水稻、矮牵牛等植物的雄性不育系中，线粒体主染色体中某些基因也决定细胞质雄性不育性的发育。例如在玉米的T型不育系线粒体中有一个叫做Urf13的自然融合基因，该基因由rrn26基因的一部分编码区和3'侧翼区以及26srRNA基因组成，由于在所有可育系线粒体基因组中都不存在这种融合基因，所以通常认为这种融合基因是造成玉米T型细胞质雄性不育的主要原因。

小　结

细胞质遗传或核外遗传特征由线粒体和叶绿体中的基因决定。不同生物的线粒体基因组变化很大，哺乳动物和酵母线粒体基因组一般为环状双链DNA，植物线粒体基因组由大环状的主染色体和小环状双链DNA或线状双链DNA组成，锥虫线粒体基因组则由多拷贝的大环和小环双链DNA组成。线粒体DNA的长度小到如莱因衣藻的4.5μm，大到如月见草的31μm。叶绿体是高等植物所特有的一种细胞器，其基因组由多拷贝的环状双链DNA组成，含有与线粒体DNA和细胞核DNA有同源性的序列。高等植物叶绿体DNA大小变动于120～160kb之间，这两种细胞器基因组都能编码各种RNA和一部分细胞器蛋白质。其他蛋白质是由核基因编码，从细胞溶质中输入的。

线粒体蛋白质基因的密码子具有不通用性，而且某些基因的信息不完整，转录后通过RNA编辑，使DNA分子中不完整的遗传信息在RNA分子上保持完整。有些叶绿体蛋白质基因为分裂基因，分成几个不同转录单位分别进行转录，然后通过反式剪接形成具翻译功能的mRNA。这两种细胞器基因的内含子都是在转录后通过自动剪接除掉的。

细胞器中有很多多亚基蛋白都是由细胞器基因和核基因共同编码的，但是细胞器基因与细胞核基因在数目上存在巨大差异，核基因表达的转录调节和细胞器基因的翻译调节是协调这些基因表达的主要机制。

线粒体和叶绿体都有各自的、与细胞核遗传体系不同的基因表达系统，但它们都必须与核遗传体系进行协调互作，才能完成各自的特殊功能，共同控制细胞的生长、繁殖和发育。

复　习　题

选择 1～3 题中正确答案：

1. GⅡ内含子剪接与细胞核前体 mRNA 剪接相似的方面有：

A. 两类内含子都有相似的剪接位点。

B. 二者都通过一种相同的、与形成套环结构有关的机制进行剪接。

C. 二者都由 RNA 所催化。

D. 二者都需要 UsnRNA。

E. 在不存在任何蛋白质的条件下，二者都能剪接。

2. 在大多数情况下，剪接过程只发生在一条 RNA 分子之内，但是 RNA 可以进行反式剪接：

A. 反式剪接涉及 2 条前体 mRNA 之间的剪接。

B. 使用一种完全不同于顺式剪接的机制。

C. 需要各种 snRNP。

3. 在锥虫线粒体中，RNA 编辑：

A. 用于全面改变某些 mRNA 的编码功能。

B. 只在每一条 mRNA 分子中发生一次变化。

C. 插入或缺失碱基 G。

D. 利用 gRNA 与前体 mRNA 中待编辑区域进行碱基配对。

4. 根据表 5－1 中资料，将糖酵母、人体和玉米线粒体 DNA 的分子量换算成碱基对数。

5. 植物线粒体 DNA 中含有哪些类型的叶绿体 DNA？为什么说线粒体中的叶绿体基因不具有功能？

6. 哺乳动物、酵母、植物和锥虫线粒体基因组中有哪些相同蛋白质基因？

7. 简述哺乳动物线粒体 DNA 复制的大致过程。

8. gRNA 与前体 mRNA 进行碱基配对有什么特点？

9. 锥虫线粒体 RNA 编辑表现什么特点？

10. 下列说法是否正确？

A. 线粒体和叶绿体 RNA 聚合酶都由各自细胞器基因编码。

B. 所有生物的线粒体 mRNA 都有 5'帽子和多聚 A 尾巴。

C. 哺乳动物线粒体的大多数基因都位于重链上。

D. 大多数叶绿体蛋白质基因的转录和翻译控制信号与原核生物基因相同。

11. 烟草叶绿体 DNA 上有一对颠倒重复，其中的基因是否都位于 DNA 的同一条单链上？理由是什么？

12. 根据表 5－7 知道，叶绿体基因组只能编码 30 种 tRNA，如果考虑到反密码子识别密码子的摆动机制，阅读完 61 种密码至少需要多少种 tRNA？

13. 简述运输肽（前导肽）在细胞溶质蛋白向细胞器运输过程中的作用。

14. 在酵母中，环境信号氧气和葡萄糖是如何调节核基因和线粒体基因表达的？

第六章

病毒基因组的表达及病毒感染对宿主细胞的遗传效应

病毒（virus）是细胞中一类复合大分子寄生物，只有在寄主细胞中才能复制。病毒的复制由自身基因型所决定，但必须利用宿主细胞的能量代谢和核酸蛋白质合成机制。病毒粒子只能含有一种核酸，要么是DNA，要么就是RNA，二者不能同时存在于同一病毒粒子之中。病毒能够自我复制，产生连续的后代，所以它们又类似生物体。但是，所有生物都是以细胞作为基本结构单位的，而病毒只是由外壳蛋白将其基因组DNA或RNA包裹成一种比宿主细胞小得多的粒子，不具有细胞特征。由于病毒颗粒微小，组成独特，并具有复制能力等原因，所以它是遗传学研究的极好材料。

第一节　病毒基因组的组织

各种病毒的基因组可分成4大类，即双链DNA、单链DNA、双链RNA和单链RNA基因组。

一、双链DNA噬菌体基因组

（一）烈性双链DNA噬菌体

1. T偶列噬菌体（T even bacteriophage）　在双链DNA噬菌体中，开头以T命名的噬菌体（从T1到T7）是一类研究非常详细的细菌病毒。T噬菌体可分成两组，一组为奇数噬菌体如T1、T3、T7等；另一组为T偶列噬菌体如T2、T4等。其中对T偶列噬菌体的研究尤为深入。

同其宿主细菌基因组一样，T偶列噬菌体为环状双链DNA基因组，具有末端重复序列。末端重复序列的长度从数个碱基对到数千个碱基对不等。这组噬菌体基因组DNA可通过脱氧胞嘧啶核苷到羟甲基胞嘧啶核苷的转换以及α-和β-葡萄糖与龙胆二糖的各种替换而高度地受到修饰。这种转葡萄糖基化发生的程度和产生的类型与噬菌体的种类密切相关。T偶列噬菌体的基因组很

大，如 T4 的基因组 DNA 长 166kb，含有 60 个以上基因。

根据 T 偶列噬菌体基因组不同部分转录时间的先后顺序，其基因组可分成早期和晚期两个区。早期区是指噬菌进入宿主细胞后直接转录的区域，晚期区则是指病毒 RNA 聚合酶合成以后，或基因组开始复制以后才转录的区域。早期区还可进一步划分“早早期”和“晚早期”区。在 T4 中，在早、晚期区之间还可划分出一个中期区。

2．T 奇数噬菌体（T odd bacteriophage）　在 T 奇数噬菌体中，T7 是研究较清楚的一种噬菌体，其基因组 DNA 长 39 936bp，含有 53 个基因。与 T 偶列噬菌体不同的是，T7 的基因组为双链线状分子，不含任何形式的修饰碱基。T7 基因组的各个顺反子之间的间隔序列很短，其 92％的基因组都用于编码（图 6－1）。

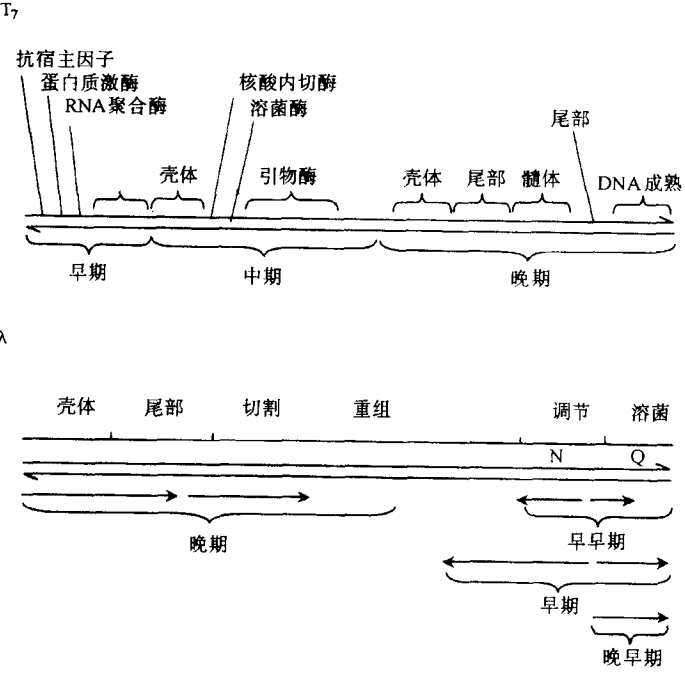


图 6－1　几种双链 DNA 噬菌体基因组的结构

（根据 L.S.Dillon 1987）

T7 的基因组可分成早期、中期和晚期三部分。早期区有 10 个基因，其产物的功能是克服宿主细胞对其入侵的限制作用，包括 RNA 聚合酶、蛋白质激酶以及一些与复制有关的酶类的基因。中间区最大，载有 22 个基因，编码 DNA 结合蛋白、内切和外切核酸酶、溶菌酶、引物酶、DNA 聚合酶以及若干种未知功能的蛋白质。晚期区有 21 个基因，大多数编码壳体和尾部蛋白。

（二）温和性双链 DNA 噬菌体

1. λ 噬菌体（λ bacterio phage） 前述 T 噬菌体在侵入大肠杆菌后，可以很快裂解细菌细胞，所以叫做烈性噬菌体。但是有些噬菌体侵入宿主细胞后，并不一定马上裂解宿主细胞，其基因组与缩主 DNA 结合，并随宿主 DNA 一起复制。

λ 噬菌体是最为熟知的温和性噬菌体之一，包裹在病毒粒子中的基因组为线状双链 DNA 分子，编码 30～40 种蛋白质。线状分子的两端具有 12 个碱基的单链尾巴，所以 λDNA 可以形成环状结构。λ 噬菌体在宿主细胞中的复制型为环状 DNA 分子，类似细菌的质粒，当噬菌体基因组复制之后，宿主细胞裂解，释放烈性 λ 噬菌体。当它侵入另一细胞后，烈性 λ 噬菌体既可以通过裂解循环进行繁殖，也可以整合到细菌染色体上，成为原噬菌体。噬菌体的裂解周期或溶原性周期的转变受 λDNA 编码的一种抑制蛋白的部分支配。λ 噬菌体基因组的突出特点是在早期阶段两条链都能转录。其基因组的早期区也可分成早早期和晚早期两部分，但这两部分转录分别受不同启动子控制。早早期基因编码几种酶，包括早期阶段所需的蛋白质 N。同样，晚早期基因的转录也是朝两个方向进行，编码与调节细胞裂解以及基因组切割和重组有关的功能，包括晚期区转录所必需的 Q 蛋白。早早期和晚早期这两个亚区位于基因组的富含 AT 的右半部分，而晚期区则主要局限在富含碱基 GC 的部分，其转录是单方向的。

λ 噬菌体整合到宿主细胞基因组中具有位点特异性。病毒基因组中与插入有关的区域叫做 attP，长 240bp，宿主基因组中插入位点叫做 attB，只有 25bp。整合过程与两种酶有关，一种叫做整合酶（integrase，Int），由病毒基因编码，能够识别细菌和噬菌体基因组中 att 位点。另一种叫做宿主整合因子（integration host factor，Ihf），由细菌基因组编码。整合的详细过程目前还不很清楚。由于大肠杆菌基因组中只有一个 att 位点，所以噬菌体始终只整合到细菌染色体的同一位点上（图 6－2）。因此，λ 原噬菌体始终位于宿主细胞的某些特殊基因之间。

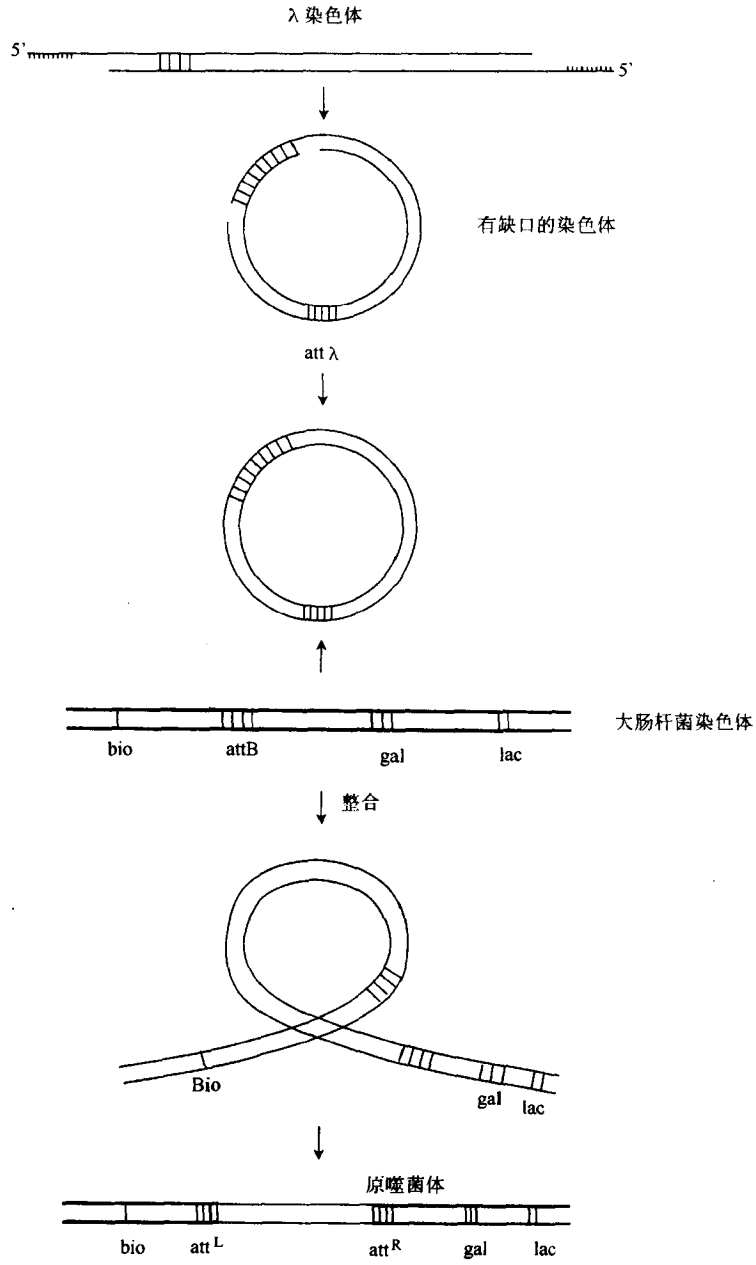


图 6－2　λ噬菌体在大肠杆菌基因组中的整合

双链的λ噬菌体 DNA 在整合到大肠杆菌染色体上之前，自发地形成一个有缺口的环

λ噬菌体从溶原性周期进入裂解周期必须从细菌染色体上切割下来，形成复制型DNA。切割过程由切割酶催化，即通过重组形成环状DNA。如果切割过程不很精确，形成的环状λDNA通常存在缺陷，可能其中某些重要的噬菌体基因被一些短的细菌DNA片段所替代（图6－3）。

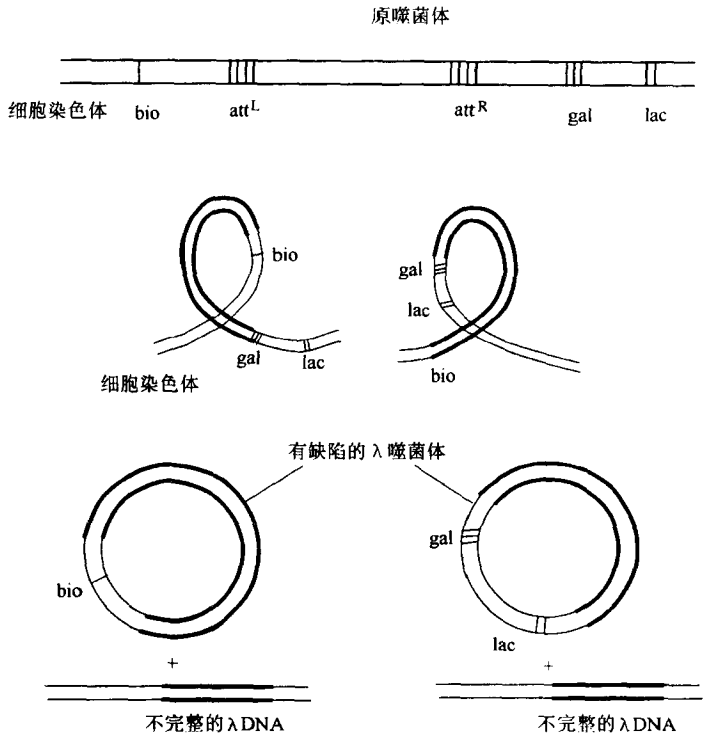


图6－3　λ噬菌体始终整合到宿主染色体的一个特殊位点上，如果切割发生差错，就会形成不完整的带有宿主基因（gal或Bio）的λ基因组

粗线示λDNA，细线示细菌染色体

2．其他温和性噬菌体　除了λ噬菌体以外，大肠杆菌还有其他一些双链DNA温和性噬菌体，目前了解较清楚的主要为P噬菌体如P1、P2和P22等。对P2噬菌体而言，其DNA分子与λ基因组有些相似，可分成早、晚期区两部分，二者转录由不同启动子控制。P22的宿主则是鼠沙门氏杆菌（*Salmonella*），其基因组也与λ相似，转录过程见后。

另一种温种性噬菌体是枯草杆菌（*Bacillus subtilis*）的噬菌体φ29，其基因组为双链线状DNA分子，长18 000bp，具有2套早期基因（图6－1C）。所有早期基因都位于一条链上，而10个晚期基因则位于另一条链上，所以二者

的转录过程具有很强的极性。另一组相近的噬菌体 M2 也侵染枯草杆菌，其详细情况以后作讨论。

二、哺乳动物双链 DNA 病毒基因组

真核生物病毒极其复杂，常根据病毒核酸类型、有无包膜、病毒基因组是单链或双链、正链或负链、分节段或不分节段、复制特点等进行分类排序，可分为以下几种类型：1. 双链 DNA 有包膜病毒（痘病毒、疱疹病毒）；2. 双链 DNA 无包膜病毒（虹彩病毒、腺病毒、乳多空病毒）；3. 单链 DNA 无包膜病毒（细小病毒）；4. 不完全双链 DNA 有包膜病毒（嗜肝 DNA 病毒）；5. 双链 RNA 有包膜病毒；6. 双链 RNA 无包膜病毒（呼肠孤病毒）；7. 正链 RNA 有包膜病毒（披膜病毒、黄病毒、日冕病毒）；8. 正链 RNA 无包膜病毒（微核糖核酸病毒、嵌杯样病毒）；9. 不分节段负链 RNA 有包膜病毒（副黏病毒、弹状病毒）；10. 分节段负链 RNA 有包膜病毒（正黏病毒、布尼安病毒、嵌沙样病毒）；11. 具有双链 DNA 的正链 RNA 病毒（逆转录病毒）；12. 未分类病毒（Marburg 和 Ebola 病毒、δ 肝炎病毒、中枢神经系统缓慢感染因子。这里主要以 4 种病毒为例进行讨论，即多瘤病毒（polyomavirus）、腺病毒（adenovirus）、疱疹病毒（herpesvirus）和痘病毒（poxvirus）。

（一）哺乳动物小分子 DNA 病毒

1. *猴肾病毒 40（SV_{40}）*　SV_{40}是多瘤病毒的代表，其基因组较小，为环状双链 DNA 分子，长 5 243bp，只有 λ 噬菌体基因组的 10%，编码 10 种分子量为 20kD 的蛋白质。DNA 分子中惟一受修饰的核苷酸为 5－甲基胞嘧啶。病毒粒子中的 DNA 与几种不同的组蛋白结合，组蛋白是以核小体的形式与病毒 DNA 连接在一起的，因此具有类似宿主细胞染色质的结构。已经证明，与病毒 DNA 相连的组蛋白为细胞组蛋白 H_{2A1}、H_{2A2}、H_{2A}、H_3。

SV_{40} DNA 基因组上有单一的 EcoRI 位点，以此位点作为物理图谱的 0 点（或 1.0 点），则 DNA 合成的启始位点位于 0.67 单位处，断开后按逆时针方向（向左）至 0.17 处为早期区，编码大 T 抗原和小 t 抗原；按顺时针方向（向右）编码结构蛋白 VP_1、VP_2、VP_3。在早期基因和晚期基因 5′端之间还有一段调节区，这段序列调节早、晚期基因的转录，同时也调节其 DNA 复制，其中含有复制启始信号（ori）。

2. *腺病毒基因组*　腺病毒基因组的大小约为 SV_{40}基因组的 6 倍，为线状双链 DNA 分子，其中腺病毒 5（Ad5）的 DNA 长 36kb。在每一条单链的末端

都有一段长约 100bp 的序列与另一端的序列具有互补性，即一种末端颠倒重复结构。所以当 DNA 变性之后，单链可自动形成一种锅柄形状。病毒粒子中的 DNA 在每一条单链的 5′端都有一种 55kD 的蛋白质（叫末端蛋白）与其共价结合。如图 6－4B 所示，腺病毒的早、晚期基因都是重叠基因，所以其基因组不像 SV_{40} 基因组可以明确地划分成早、晚期区。早期基因（从 E1 到 E4）分散在基因组的不同区域（图 6－4B）。它们位于两条不同单链上。每一区域都可产生多种转录子，同一区域的各种转录子都是前体转录子通过可变剪接途径形成的，它们都保留了前体 RNA 的 5′ 和 3′末端，只是其内部区域不同。在某些情况下还可产生重叠 mRNA，即由同一可读框编码的相似但不完全相同的蛋白质。有些 mRNA 还可利用不同的重叠可读框编码出氨基末端部分相同、而羧基末端部分不同的蛋白质。晚期基因可分成 5 个区域（从 L1 到 L5），它们编码约 12 种病毒粒子的壳体蛋白，其转录可能都受同一启动子（P_L）控制。这个活性极强的启动子指导合成一条很大的初级转录子，通过转录后加工，至少可产生 20 种不同的 mRNA。所有这些晚期基因都由宿主细胞的 RNA 聚合酶Ⅱ转录，转录子都具有 5′帽子结构和 3′多聚 A 尾巴。但另一叫做 VA 的基因则由宿主细胞的 RNA 聚合酶Ⅲ转录，其启动子也位于转录区之中，这种 RNA 与晚期 mRNA 的翻译有关，其功能可能是将晚期 mRNA 运输到宿主细胞质中。

（二）脊椎动物大双链 DNA 病毒

有些椎脊动物病毒除了具有很大的基因组以外，还具有包裹壳体的脂质外膜，其基因组为线状双链 DNA。这组病毒中研究较清楚的有疱疹病毒和痘病毒。

1. *疱疹病毒基因组* 疱疹病毒包括疮疹病毒 1（HSVI）、疱疹病毒 2（HSV2）和 EB 病毒（Epstein Barr Virus），为一些致癌的可疑因子（见本章第四节）。这 3 种病毒的 DNA 都绕蛋白质核心形成超螺旋。HSV1 和 HSV2 基因组长 150kb，EB 病毒基因组则长 172kb。总的说来，这三种病毒的基因组都具有 5 个特殊区域（从 U1 到 U5），各个区域之间都由 4 个内部重复序列将其隔开，基因组的两端都具有末端重复。在病毒侵染循环的后期，基因组转录后形成 3 条多顺反子 mRNA（图 6－4C），其中一条多顺反子 mRNA 的转录方向与另外两条多顺反子 mRNA 的方向相反。在 U3 区域中还存在 2 个早期基因，由于这两个基因的转录方向与 U3 的转录方向相反，所以二者位于晚期基因所在链的互补链上。

2. *痘病毒基因组* 痘病毒中具有致命性的病毒是天花病毒（small pox virus），但研究最清楚的是牛痘病毒（vaccinia virus）。牛痘病毒是痘病毒的原

始类型，其基因组比 EB 病毒基因组略大，长 187kb。痘病毒基因组最明显的特征是 DNA 一端的两条单链末端通过磷酸二酯键共价结合在一起，所以当 DNA 变性后就可形成一个连续的单链 环状分子。痘病毒 DNA 分子的第二个特点是有长短不等的末端重复，每一条单链的两个末端都具有互补性，有些病毒的末端重复可长达 10kb，痘病毒基因组至少有 3 个在感染初期转录的早期基因。其基因组表达的特点之一是早期基因在壳体内转录，晚期基因在缩主细胞质中转录。早期基因和晚期基因紧紧相连，其中有些基因还相互重叠。据估计，痘病毒基因组至少含有 150 多个基因。

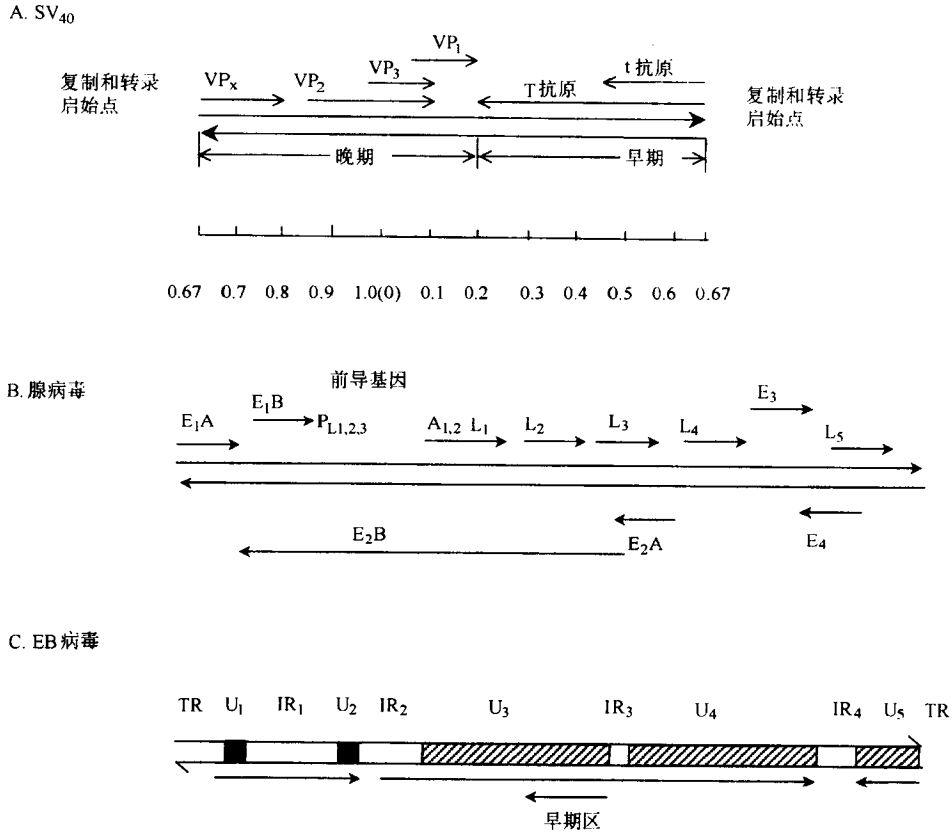


图 6－4　哺乳动物双链 DNA 病毒的基因组

A. 环状 SV40DNA 分子以 EcoRI 的单一酶切位点作为 0（或 1.0）点，转录启始点则在 0.67 图谱单位处，断开后以线形表示　B. 腺病毒的早期基因分散在两条不同单链上，P_L 为晚期基因的主要启动子　C. 疱症病毒的 EB 病毒，其基因组分为 5 个特殊区域（U1－U5）和 4 个内部重复序列（IR1～IR4），其末端为末端重复（TR）

三、植物双链 DNA 病毒基因组

在高等植物中，目前已知的唯一双链 DNA 病毒为花椰菜花叶病毒（caulimovirus，CaMV），其基因组为环状双链 DNA 分子，长 8 024bp，具有 7 个编码区，都位于负链上（图 6－5）。每两个相邻编码区之间都有一段短的间隔序列，其中基因Ⅵ和基因Ⅶ之间的间隔序列长 700bp。CaMV 基因组有 3 个断裂区，第一个断裂区在基因Ⅵ和基因Ⅶ之间，位于负链上，另外 2 个断裂区在正链上。各个断裂区上都有一段单链重复，形成一种三链结构，其功能可能与转录有关。CaMV 的复制首先以 RNA 为模板，即从断裂区 I 附近开始的整个基因组转录子 35SRNA 为模板，通过反转录，形成基因组 DNA。

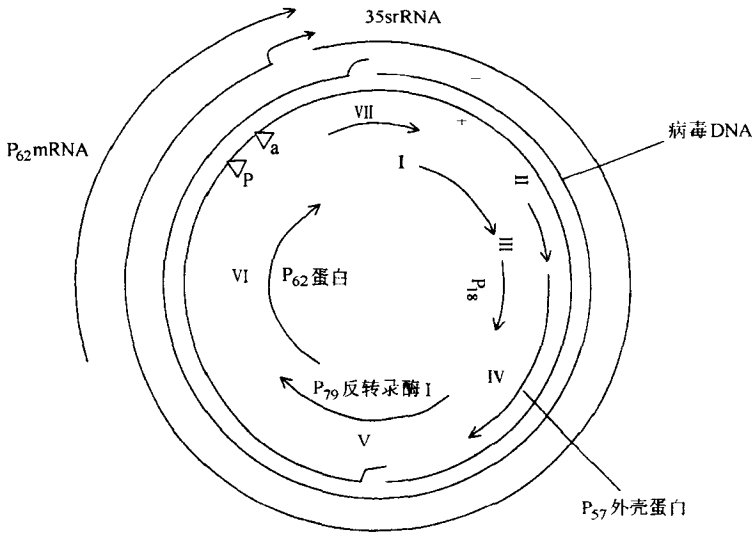


图 6－5　CaMV 的基因及其转录子

所有基因（Ⅰ－Ⅶ）都位于一条 DNA 链上，转录子 35S RNA 具有一条短的末端重复序列；P－35S RNA 的启动子，a－多聚 A 信号序列

四、单链 DNA 病毒基因组

单链 DNA 病毒主要见于原核生物，植物中也有少数几种单链 DNA 病毒。

（一）细菌单链 DNA 病毒

根据壳体的形状划分，细菌病毒可以分成两组，即廿面体形（球形）和丝

状体形。甘面体形单链 DNA 病毒主要有 φX174、M13 和 Φ1，这三种病毒的基因组均为单链环状 DNA 分子，侵染宿主细胞后，病毒首先通过合成基因组 DNA 的互补链，形成双链环状分子，即所谓复制型（replicative form）。复制型分子以两种状态存在，一种是闭合环状的超螺旋结构，另一种是松弛结构，而且其中一条单链上至少有一个缺口。例如 φX174 的单链基因组长 5 386 个碱基，含有 10 个编码蛋白质的基因，其中多数都是重叠基因。在第一个编码区即蛋白质 A 的编码区之前有一段 63nt 的间隔序列，基因 J 和 F 之间有一段 39nt 的间隔序列，基因 F 和 G 之有一段 110nt 的间隔序列，在基因 G 和 H 之间有一段 8nt 的间隔序列，而其他顺反子或者是重复基因，或者只相隔一、二个碱基。

研究较清楚的丝状单链 DNA 病毒主要有 f1、fd、M13 和 Ike，这类病毒的基因组与甘面体形病毒基因组相似，当单正链基因组 DNA 进入细菌细胞后立即复制负链，形成双链复制型分子。在这组病毒中，各种病毒基因组的大小和基因分布都非常相似，也都是环状分子（图 6－6），基因组大小变动于 5 781nt（病毒）到 6 883nt（IKe 病毒）之间。这类噬菌体在感染细胞后并不合成裂解蛋白，也不裂解细胞，而是随着宿主细胞生长，以一种温和方式合成噬菌体蛋白和 DNA，所以不会中断宿主细胞的代谢过程。

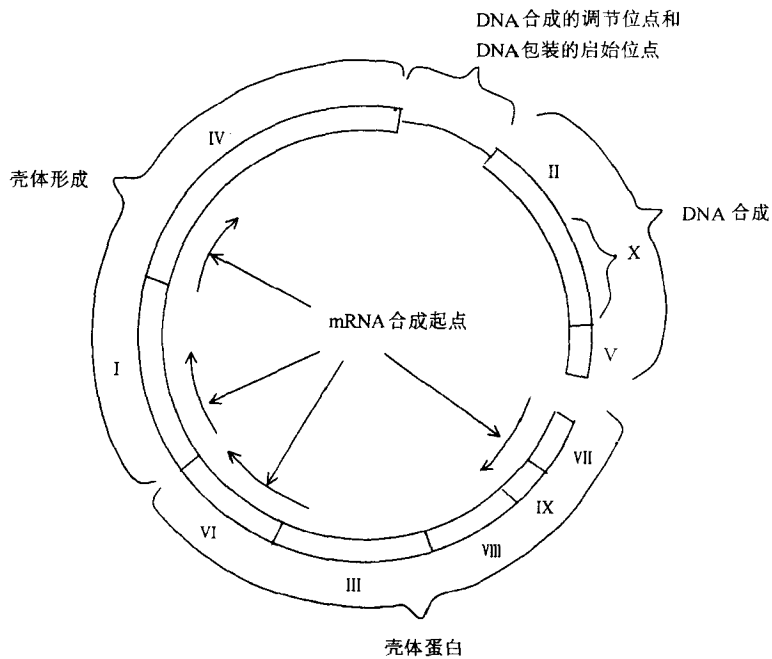


图 6－6　丝状体形单链 DNA 噬菌体（fd）遗传图

（二）植物单链 DNA 病毒

高等植物中惟一具有单链 DNA 基因组的病毒是双病毒（geminivirus）。这组病毒为双扁球形颗粒，宿主范围较广泛，包括单子叶植物和双子叶植物。根据对菜豆金色花斑病毒（bean golden mosaic virus BGMV）和木薯潜伏病毒（cassava latent virus CLV）的研究，双病毒 DNA 为单链开环分子，可分成两种不同的核酸序列，双病毒的每一个成员只能含有其中一种 DNA，所以为二基因组病毒。如 CLV 病毒含有两组不同序列的 DNA1 和 DNA2，二者分别长 2 780和 2 724nt。虽然这两种 DNA 分子绝大部分不同，但二者之间有 219 个碱基的序列存在同源性。也有某些双病毒为单基因组病毒，如玉米条纹病毒（maizestreakvirus，MSV）和虎尾草条纹花斑病毒（chloris striate mosaic virus，CSMV）只有一条单链 DNA 分子。

有关双病毒 DNA 的复制和转录还不很清楚，但知道在双病毒 DNA 中有一部分为双链 DNA，这种双链区可能是复制中间产物。另外，在 CLV 病毒的 DNA1 和 DNA2 的同源序列中，各有一段能够形成发夹环的序列，这种结构可能用作 DNA 的复制启始区。

五、RNA 病毒基因组

（一）单链 RNA 噬菌体

目前知道的 RNA 噬菌体有 4 组，其中有 2 组的分子生物学研究较清楚，其代表为 MS2 和 Qβ。所有单链 RNA 噬菌体的基因组都很小，只有 3 500 到 4 500个核苷酸，为单正链 RNA，被包裹在 20 面体的壳体之中。MS2 基因组长 3 569nt，含有 4 个基因，分别编码壳体成分 A 蛋白和外壳蛋白、裂解酶以及病毒复制酶的 β 亚基，但后二者为重叠基因（图 6－7）。病毒复制酶的其他亚基由宿主基因编码，而且 MS2 基因组的复制、转录、翻译和病毒粒子的组装都高度依赖于宿主细胞。

Qβ病毒为负链 RNA 基因组，所以在蛋白质合成过程中，需要将基因组转录成 mRNA。Qβ 的基因组长 4 220 个核苷酸，能编码复制酶（图 6－7B）。这种复制酶只能复制 QB 基因组以及其他几种与 QB 相似的噬菌体基因组。它既不能复制其他任何病毒基因组，也不能复制大肠杆菌 RNA。复制酶对模板的结构要求非常严格，它只有识别 Qβ 基因组的一段内部序列，而且只有基因组 3′端序列保持完整时才能识别这种复制启始信号。

（二）多细胞动物的单链 RNA 病毒基因组

多细胞动物的 RNA 病毒种类繁多，现以其中几种重要病毒为例对其基因

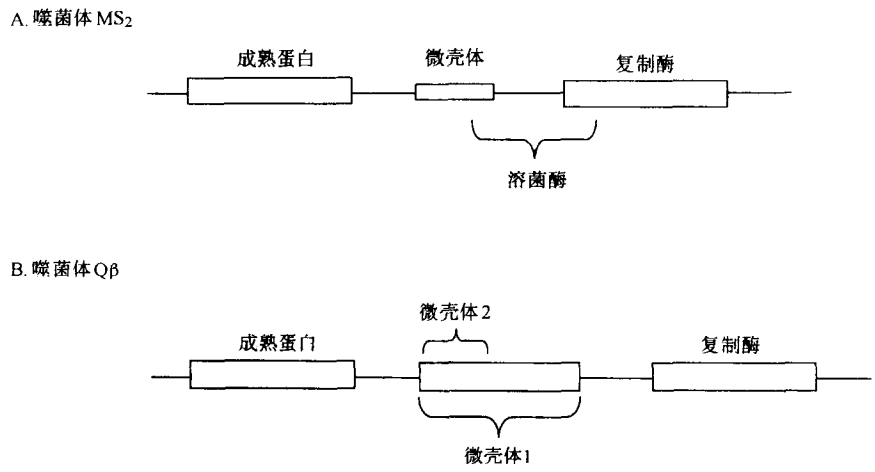


图 6－7　两种重要的单链 RNA 噬菌体的基因组的结构

组结构加以讨论。

1. 微小 RNA 病毒　微小 RNA 病毒包括感冒致病因子鼻病毒（rhinovirus）、脑炎心肌炎致病因子心病毒（cardiovirus）、动物口蹄疫致病因子口疮病毒（aphtho virus）以及脊髓灰质炎和甲型肝炎致病因子肠病毒（enterovirus）。所有这些病毒的基因组都是正链 RNA，为一种编码所有病毒结构蛋白和酶蛋白的多顺子 mRNA。例如，鼻病毒 14 的基因组长 7 212nt，只含有一个可读框，由 6 540 个核苷酸组成。甲肝病毒（hepatitis）的基因组稍长，有 7 478nt，也只有一个由 6 682nt 组成的编码区。所有这些病毒的基因组都只编码三种蛋白质。

根据对三种不同血清型的脊髓灰质炎病毒（poliovirus）的研究，其基因组长 7 500nt，比鼻病毒基因组略大，但其编码特性却复杂得多。脊髓灰质炎病毒 RNA 结构为 5′－VPg－NTR－poly（A）－3′，其中，NTR 是非常保守的非翻译区，位于多聚蛋白编码区的两侧，可能携带有复制酶的序列。多聚蛋白基因组 mRNA 翻译后只产生一条多蛋白质分子，先被水解成 P1、P2、P3 三个次级前体，再进一步被分别切成 4 个、3 个和 4 个终末产物。病毒基因组 RNA5′端接有由病毒基因编码的一种蛋白质，3′端接有多聚 A 尾巴（图 6－8）。

尽管上述各类病毒基因组的结构非常相似，但它们的 RNA 序列几乎不存在任何同源性，即使在同一类病毒如鼻病毒和脊髓灰质炎病毒中也只有几种蛋白质具有 44％～65％的序列同源性。

2. 黄病毒基因组　黄病毒（flavivirus）包括 70 多种非常相近的、对人体和家畜具有致病性的病毒，其中绝大多数由吸血昆虫和虱传播，包括脑炎病

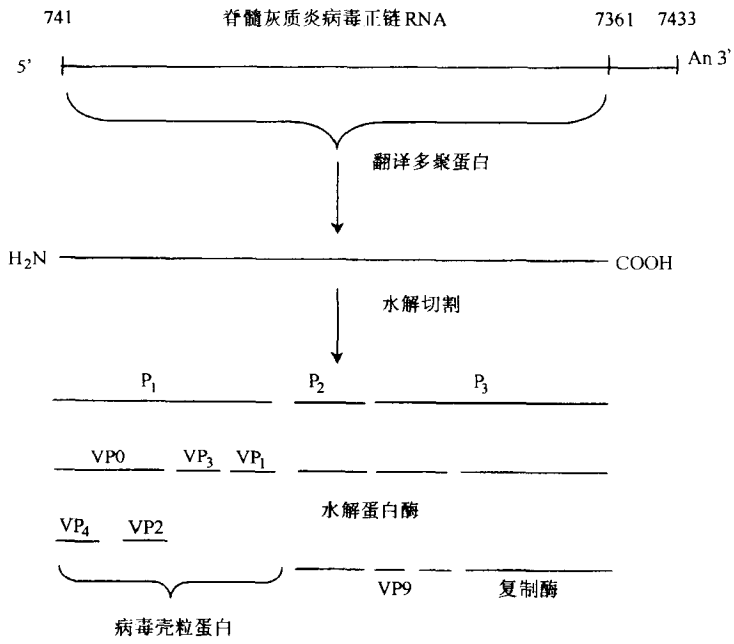


图 6-8　脊髓灰质炎病毒基因组及其翻译产物

前体多蛋白质在翻译尚未完全结束时就水解成 P1、P2 和 P3。所有切割过程都由病毒基因编码的蛋白质催化

毒、登格热病毒和黄热病毒。这些病毒的基因组也都是单正链 RNA，长 11 000nt 左右，含有 9 个基因，其中有 3 个基因编码结构蛋白，其他基因编码水解酶和一些非结构蛋白。

3. 棒病毒基因组　棒病毒（rhabdovirus）呈子弹形（图 6-9），也为单链 RNA 基因组，但为负链 RNA 分子，所以在蛋白质合成过程中需要将基因组转录成 mRNA。其中最为熟知的致命性病毒就是狂犬病毒（rabies virus），但研究最清楚的是温和性的水泡性口炎病毒（vescular somatitis virus，VSV）。水泡性口炎病毒基因组编码 6 种蛋白质，其中有一个基因位于前导序列中，只有 47 个碱基（见本章第三节）。由于整个基因组只在前导序列之前有一个启动子，所以转录具有极性和顺序性，即 5′顺反子的转录频率高，3′顺反子的转录频率低。同所有哺乳动物的 RNA 病毒一样，水泡性口炎病毒基因组也被包裹在核蛋白颗粒之中，形成核壳体（nucleocapsid）。

4. 大 RNA 基因组病毒　大 RNA 病毒又称副黏病毒（paramyxovirus），其基因组也为负链 RNA，其中最具代表性的病毒为仙台病毒（Sendaivirus）。仙台病毒基因组长 15 000 个碱基，编码 7 种用于形成壳体的结构蛋白和一种未知功能

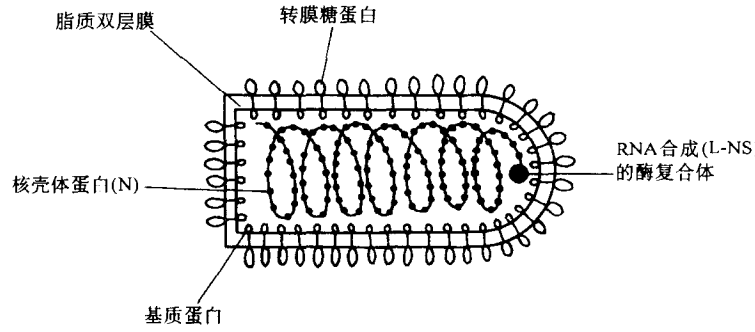


图 6－9　水泡性口炎病毒粒子
示棒病毒典型的子弹形特征

的非结构蛋白。基因组中大多数基因都有各自的启动子，有些基因高度重叠。

在多细胞动物中还存在其他一些大 RNA 病毒如流感病毒 A（influenza A virus）也是一种单负链 RNA 病毒，但其基因组由 8 条 RNA 分子组成。每一条 RNA 都编码不同蛋白质，而且都与自身的转录酶结合并被包裹到各自的呈螺旋状的核壳体之中（图 6－10）。流感病毒基因组的各条 RNA 片段的 5′端均是有 13 个核苷酸组成的相同序列，3′端前面的 12bp 序列高度保守，且两端的 15bp 序列互补，所以 RNA 分子可以环化，形成末端呈锅柄状结构。流感病毒与其他 RNA 病毒的另一主要差别就是其复制依赖于宿主细胞核。如果通过实验操作或用药物处理将细胞核除掉，流感病毒就不能在细胞中进行复制。

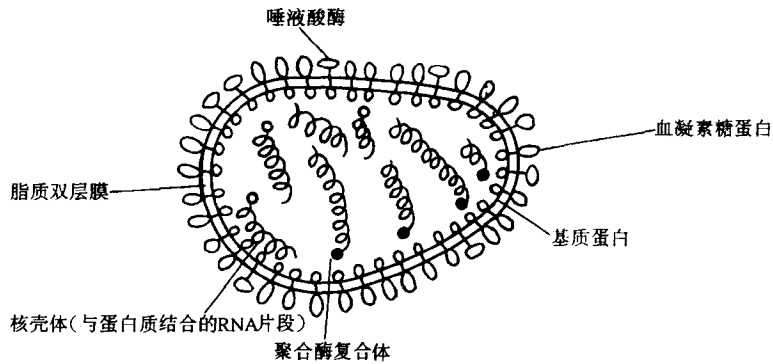


图 6－10　流感病毒粒子
示其 8 条 RNA 基因组片段

流感病毒呈卵圆形，具有两种不同的表面糖蛋白。一种为血凝素糖蛋白（hemagglutinin glycoprotein），其功能是识别细胞表面受体和使病毒在发生细

胞吞食之后进入细胞膜。另一种为唾液酸酶（neura minidase），其功能是从细胞表面蛋白上除掉唾液酸（sialic acid）基团。同 VSV 一样，流感病毒也含有基质蛋白和核壳体蛋白。

（三）植物单链 RNA 病毒

在目前已知的所有植物病毒中，90%以上都是 RNA 病毒，其中 76%为单正链 RNA 基因组，13%为单负链 RNA 基因组。如果按照病毒基因组的组成进行划分，植物 RNA 病毒可分成三组。第一组为单基因组 RNA 病毒，其中大多数成员的基因组都是单正链 RNA。第二组为二基因组病毒，其基因组通常由两条不同的 mRNA 分子组成，但二者被分别包裹在不同颗粒之中，只有二者相互合作才能感染植物。第三组为三基因组病毒，即三条单链 mRNA 分子分别包裹在三个不同颗粒中，只有这三条 RNA 分子协调合作，三基因组病毒才能感染植物。

除了基因组 RNA 以外，许多植物病毒还具有若干较小的亚基因组 mRNA（subgenomic －messenger RNA），其功能是扩增某些特殊基因产物如外壳蛋白等。亚基因组 RNA 起源于基因组 RNA，在某些病毒中，亚基因组 RNA 可以与其相应的三基因组或二基因组中一条基因组 RNA 分子共处于一个颗粒中，而在另一些病毒中，则独立地包装成一种较小的颗粒。植物 RNA 病毒中也有一组棒病毒，其特征都与动物棒病毒相似，其基因组也是一条单负链 RNA 分子，表达时病毒粒子中的转录酶以基因组 RNA 为模板，合成相应的 mRNA 分子。

大多数正链 RNA 病毒的基因组 RNA 与真核细胞的 mRNA 基本相似，其 3′端都有一段多聚 A 尾巴（图 6－11）。多聚 A 尾巴是病毒基因组的结构特征之一，它不同于细胞内 mRNA 的多聚 A 尾巴，宿主细胞 RNA 的这种尾巴是在转录后由酶加上的。另一类基因组 RNA 的 3′端不具有多聚 A 尾巴，但接有一段类似 tRNA的结构。还有一类病毒基因组的 3′端既不具有多聚 A 尾巴，也不具有类似 tRNA 的结构，只暴露出一个游离的羟基。正链基因组 RNA 的 5′端也可区分成几种不同结构，大多数 RNA 都具有帽子结构，但也有一些病毒的基因组 RNA 只有 5′末端二磷酸或三磷酸基团。有些病毒基因组的帽子结构往往与一种叫做 VpG 的蛋白质共价结合，这种蛋白质可能与 RNA 复制有关。

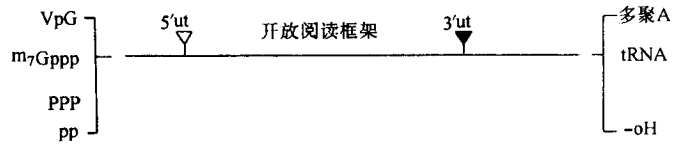


图 6－11　正链 RNA 病毒基因组的一般结构

启点密码，－终止密码，ut－非常编码区

（四）卫星病毒基因组

高等真核生物的某些 DNA 和 RNA 病毒还具有卫星病毒（satellite virus）或卫星 RNA，如烟草坏死病毒（tobacco necrasisvirus，TNV）的卫星病毒为一条包裹在不同颗粒中的 mRNA 分子，编码卫星病毒的外壳蛋白。TNV 卫星病毒的复制完全依赖于 TNV 帮助病毒（helper virus）。卫星病毒与帮助病毒的 RNA 序列不存在同源性，如黄瓜花叶病毒（cucumber mosaic virus，CMV）与其卫星病毒的 RNA 序列就不相同，尽管如此，但卫星 RNA 的复制与表达则完全依赖于 CMV 帮助病毒，并且与帮助病毒基因组共处于一个颗粒中。卫星病毒或卫星 RNA 可以影响由帮助病毒所导致的植物病症，干扰帮助病毒的复制过程，所以卫星病毒或卫星 RNA 可能是帮助病毒的一种分子寄生物。

除了植物的某些 RNA 病毒具有卫星病毒以外，哺乳动物和人体的某些 DNA 病毒也都有卫星病毒。例如，人体乙型肝炎病毒（hepatitis B virus）有一种卫星病毒叫做 δ 肝炎病毒（hepatitis delta virus，HDV）。HDV 的基因组为环状单链 RNA，长 17 000nt，其中 70% 的碱基可以进行互补配对，所以能够折叠成棒状结构。其基因组 RNA 可以分成两部分，第一部分为非编码部分，但其中存在与基因组复制有关的顺作用子序列。第二部分为蛋白质编码部分，已知其中有一个 δ 抗原（195 个氨基酸）的顺反子。在受感染的肝细胞中存在 2 种主要的 HDV 的 RNA（图 6－12），第一种 RNA 叫做反基因组（antigenome），同基因组 RNA 等长；第二种 RNA 叫做互补 RNA，比基因组 RNA 略短，它是 δ 抗原的 mRNA。反基因组 RNA 存在于受感染细胞的细胞核中，而 mRNA 则存在于细胞质中。反基因组 RNA 同基因组 RNA 一样，也可折叠成棒状结构。基因组 RNA 和反基因组 RNA 都含有自动切割位点，自动切割后的线状分子可以通过形成棒状结构而将两端自动连接在一起。关于反基因组 RNA 形成的机制目前还不清楚。

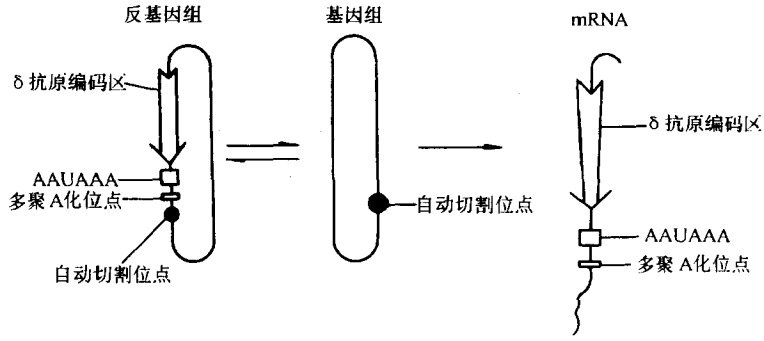


图 6－12　HDV 基因复制的三种 RNA

这种卫星病毒的复制依赖于帮助病毒（乙型肝炎病毒），在乙型肝病毒（见本章第二节）的帮助下，HDV 在肝细胞中高效复制，并增加由乙型肝炎病毒感染所引起的肝损坏的程度。在组装成病毒粒子时，HDV 也必须利用帮助病毒的外壳蛋白。

（五）双链 RNA 病毒基因组

真核细胞和原核细胞都有基因组为双链 RNA 的病毒。在真核生物病毒中，有三组病毒含有多条双链 RNA 分子，即桑蚕的细胞质多角体病毒（cytoplasmic polyhedrosis virus）、哺乳动物的呼肠孤病毒（reovirus）和轮状病毒（rotavirus）。前两组病毒的基因组包括 10 条双链 RNA，后一组病毒有 11 条 RNA。在所有这些病毒的壳体中都贮存有由病毒基因编码的转录酶，而且每一条双链 RNA 片段都编码不同蛋白质。

对呼肠孤病毒而言，病毒粒子通过细胞吞噬作用（endocytosis）进入细胞，到达细胞核附近。病毒粒子在通过吞噬小泡期间，约有一半壳体蛋白被溶菌水解酶除掉，使病毒粒子中的转录酶被活化，进行转录并释放出新合成的 mRNA。每一条基因组 RNA 都有一条正链 mRNA，而且每一条 mRNA 的 5′端都有一个帽子结构，但不具有多聚 A 尾巴。所以新合成的 mRNA 既可用作翻译的模板，也可用作基因组复制的模板。在感染的后期，新合成的壳体蛋白同正链 RNA 结合，组装成叫做前髓部（precore）的类似壳体的结构。在前髓部内，病毒粒子的聚合酶将单正链 RNA 转换成双链 RNA 分子。前髓部用于合成新的 mRNA，使复制过程继续进行，直到细胞内形成大片复制病毒区时，复制过程才停止。在感染的末期，前髓部通过加上外壳蛋白转变成新的病毒粒子。

六、逆转录病毒基因组

逆转录病毒（retrovirus）是自然界广泛存在的一种单正链 RNA 病毒，有数百种变异体。根据其生物学特征，形态特征以及基因组结构，可将其分成三组。第一组为癌病毒（oncovirus），它是爬行动物、鱼类、鸟类、啮齿动物、猫科动物以及灵长目动物的与癌变有关的逆转录病毒，如劳氏肉瘤病毒（Rous sarcomavirus）、禽类白血病毒（avian leukosis virus）、鼠白血病毒（murine leukemia virus）、小鼠乳腺瘤病毒（mouse mammary tumor virus）以及人体 T 细胞淋巴瘤病毒（human T cell lymphotropic virus）等。第二组为慢病毒（lentivirus），损伤神经和诱发慢性肺炎，如绵羊髓鞘脱落病毒（visna virus）和山羊脑炎－关节炎病毒（caprine encephalitis－arthritis virus）以及艾滋病毒（human immunodefeciency virus，HIV）等。第三组为泡沫病毒（spumavirus），

能在几种哺乳动物不出现病症的情况下进行持续感染。

逆转录病毒基因组为二倍体，每一病毒粒子都含有两个相同拷贝的、7～10kb 的基因组 RNA。形成二倍体基因组的原因不清，可能二倍体对复制期间基因组损坏的修复以及高频率的重组具有某种促进作用。

逆转录病毒基因组在 5′和 3′端都有相当长的非翻译区，其中含有复制和翻译所需的信号序列，两端的一段短的顺式重复（R）以及通过碱基配对与基因组 5′端结合的 tRNA 分子（图 6－13）。基因组中与 tRNA3′部分互补的序列称为引物结合位点（PB）。tRNA 是病毒基因组在复制期间从宿主细胞获得的，是 DNA 合成的引物。

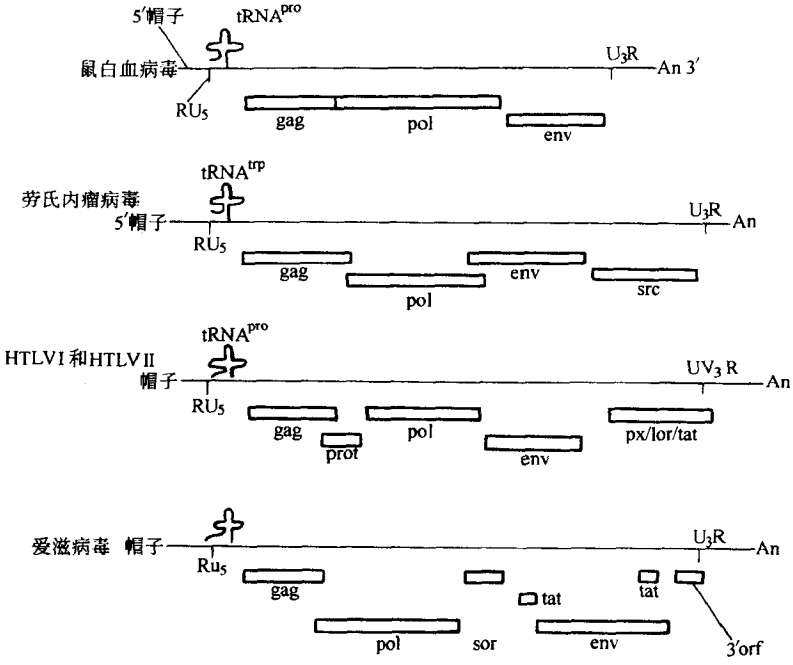


图 6－13　部分逆转录病毒基因组

长条示可读框

在爱滋病毒中 tat 基因的可读框含有一个内含子，大多数逆转录病毒的基因组都与鼠白血病毒基因组的结构相似，只有 gag、pol 和 env 基因，编码 env、src、tat 以及 sor 和 orf 的 mRNA 可能通过 RNA 可变剪接途径形成

所有逆转录病毒都具有一个编码 逆转录酶（reverse transcriptase）的基因（pol 基因）。逆转录酶为一种依赖于 RNA 的 DNA 聚合酶，可将病毒基因组 RNA 逆转录成线状双链 DNA，叫做原病毒（provirus）。这种双链 DNA 比基因组 RNA 稍长，因为其两端接有 300～1 000bp 的特殊的长末端重复序列

(long terminal repeat, LTR)。LTR 起源于基因组 RNA 两端的序列，但病毒粒子中的基因组 RNA 并不含有 LTR。逆转录过程依赖于病毒基因组 3′端的多聚 A 序列和近 5′端的一种 tRNA 序列。逆转录的详细过程现在还不很清楚。双链 DNA 分子合成后就整合到宿主细胞基因组中，随宿主基因一起遗传和表达。

致癌性逆转录病毒也可进一步分成两组，一组为复制胜任型（replication compe tent），其基因组缺少起源于宿主细胞的癌基因，但具有与基因组复制有关的所有基因。另一组为复制缺陷型，也称急性转化逆转录病毒，这组病毒含有衍生于宿主细胞的癌基因，但缺少某些与基因组复制有关的基因。这两组病毒在诱发恶性肿瘤的方式方面也有所不同。缺陷型逆转录病毒诱发肿瘤形成的时间很短，通常只在几星期内就可使宿主细胞形成急性肿瘤。复制胜任型病毒一般诱发慢性恶性肿瘤，潜伏期较长，在感染后 3 至 12 个月内，细胞不会恶性生长。由复制胜任型病毒诱发的常见病变是白血病（leukemia），由复制缺陷型病毒诱发的恶性肿瘤通常为肉瘤（sarcoma）和癌（carcinoma），劳氏肉瘤病毒就属于这一类型。

过去曾认为复制胜任型逆转录病毒只有 3 个基因即 gag、pol 和 env 基因。根据对三种人体 T 细胞白血病毒（HTLV）的研究，逆转录病毒基因组中还有一个与 gag 和 pol 序列重叠的编码核酸酶的顺反子。此外，env 基因后的 X（或 R）区中也存在一个可读框，可能这段序列也具有编码功能。

gag 和 pol 基因是研究较清楚的两个基因，二者都编码多蛋白质分子，为一种组合蛋白基因。同隐密蛋白基因相比，组合蛋白基因在翻译后，迅速将其编码产物加工成各种不同成分，其各个蛋白质部分并不以隐密形式贮存在细胞中。gag 基因编码分子量分别为 15kd、12kd、30kd 和 10kd（从 5′→3′）的四种多肽，它们都存在于病毒粒子的髓部，其中 30kd 的蛋白与病毒识别宿主细胞有关。env 基因也是组合蛋白基因，编码两种组成壳体的蛋白质，翻译后的初级蛋白在糖基化后再切割成两条多肽（gp70 和 p15E）。在摩氏白血病毒中，还有一个叫做 R 的基因与 env 顺反子重叠。在 HTLV－Ⅰ和Ⅱ病毒的 3′末端区中则分别有一个编码分子量为 41kd 和 38kd 的蛋白质的顺反子。

根据对各种复制缺陷型病毒的研究，其基因组中都有一个与宿主基因存在同源性的病毒癌基因，叫做 V－onco 基因。这些病毒癌基因的名称因病毒种类不同、发病动物不同以及编码产物和肿瘤类型不同，名称各异，如劳氏肉瘤病毒的癌基因叫做 V－src，猴肉瘤病毒（simian sarcoma vius）的癌基因叫做 V－sis 等。根据 DNA 分子杂交证明，鸡的正常细胞和癌细胞含有与劳氏肉瘤病毒的 src 基因存在同源性的基因，细胞内这种与 src 有关的基因叫做 c－src。细胞内的 c－src 基因为正常基因，但具有转变成癌基因的可能性，所以叫做原

癌基因（proto oncogene）。而劳氏肉瘤病毒中的 V－src 癌基因可能是通过某种类似噬菌体转导的方式从细胞中获得的癌基因。

七、植物类病毒和拟病毒基因组

植物中有一种基因组非常特殊的致病因子叫做类病毒（viroid）。类病毒为一种裸露的 RNA 分子，长 240～350 个核苷酸，为单链环状结构，其分子内许多区域都可进行碱基配对，所以呈一种棒状、近于双螺旋的结构。类病毒基因组不编码任何蛋白质，如马铃薯块根类病毒（patato spindle tuber viroid）基因组不含 AUG 启点密码，所以类病毒 RNA 不编码自身基因组复制所需的蛋白质，其复制完全依赖于宿主植物细胞的 RNA 聚合酶Ⅱ。类病毒 RNA 基因组按滚环复制模型进行复制。

在植物某些 RNA 病毒粒子中存在一种叫做拟病毒（virosoid）的 RNA 分子。拟病毒 RNA 的结构与类病毒 RNA 的结构相似，但二者之间并不存在任何序列同源性。拟类病毒不具有编码功能，也不能独立地复制，其复制过程依赖于帮助病毒。另一方面，其帮助病毒的复制过程也需要拟病毒，如果缺少拟病毒，帮助病毒基因组 RNA 就不能复制。

第二节　病毒基因组的复制

一、线状双链 DNA 病毒基因组的复制

（一）T_7 噬菌体基因组的复制

T_7 噬菌体基因组复制所需的各种蛋白质都由自身基因组编码，这些蛋白质具有复制双链 DNA 的基本活性，但与大肠杆菌的 30 种复制酶或蛋白质相比，其数目少得多。由基因 5 编码的 DNA 聚合酶在与细菌基因编码的一种硫氧还原蛋白（thioredoxin）形成复合体后才具有活性，基因 4 编码的蛋白质具有引物酶和 DNA 解螺旋酶的作用，基因 1 和 3 编码 DNA 连接酶、基因 2 编码单链 DNA 结合蛋白，基因 6 编码除掉引物 RNA 的核酸酶。这些蛋白质就是 T_7 基因组复制所需的所有酶类。

在感染宿主细胞的过程中，不同 T_7DNA 分子之间进行交换所需的蛋白质也由 T_7 基因组编码，但重组过程需要基因 6 编码的核酸酶和另一种由基因 3 编码的核酸酶。可能这两种酶通过局部降解正在复制的 DNA 分子，启始杂二

聚体形成（第一章）。

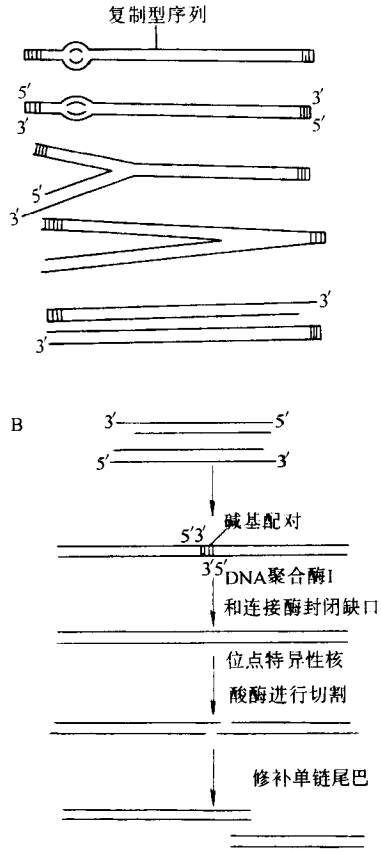


图 6－14　T_7DNA 分子的复制

A. 线状 DNA 的 3′端首先进行不完全复制　B. 两条不完整的双链接在一起，然后 DNA 聚合酶按 5′→3′方向将二聚体复制完整

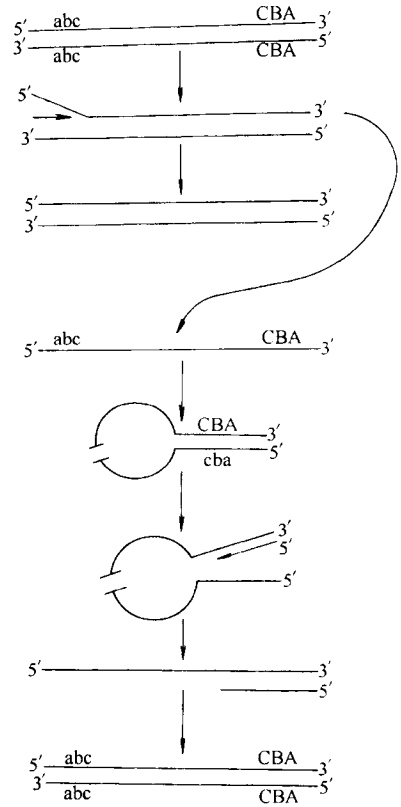


图 6－15　线状 DNA 分子

（如腺病毒和噬菌体 φ29）复制时并不使用引物 RNA，复制可从 DNA 的任一端开始，其中只有一条链用作模板，被置换的单链脱离 DNA 亲链，并通过两端的颠倒重复进行互补配对，形成发夹环，然后再以这种结构作为模板，从 3′端开始合成新的子链

如前所述，T_7 基因组为线状双链 DNA，其两端都有长约 160bp 的额外重复序列。但是这种线状 DNA 分子复制时并不产生单个基因组单位，而是形成一段很长的、由多个基因组单位首尾衔接在一起的复合基因组（图 6－14）。形

成这种多基因组 DNA 的原因是由于线状双螺旋在第一次复制时产生了两条不完整的子代分子。当每一条子双螺旋的一端完整地复制时，另一端则留下一段 3′单链尾巴。这两条双螺旋的单链尾巴可进行碱基互补配对，当填补了缺口以后，就产生含有 2 个基因组单位的 DNA 分子。这种二基因组 DNA 然后又通过复制，将两条子双螺旋再连接在一起，形成一条含有 4 个基因组单位的分子，并如此继续下去。然后，这种多基因组单位的 DNA 通过特殊内切核酸酶切割，形成具有黏性末端的单个基因组分子。最后，DNA 聚合酶按 5′→3′方向填补单链尾巴，形成齐头末端，产生与原来亲本噬菌体基因组完全相同的 DNA 分子。

（二）腺病毒基因组的复制

某些原核细胞和真核细胞病毒的线状双链 DNA 基因组在复制时并不需要合成引物 RNA，而且合成过程是连续的，并不形成冈崎片段。如腺病毒以及枯草杆菌噬菌体 φ29 等。这类病毒基因组的复制都无双向复制叉，复制过程只从线状分子的一端或另一端开始。虽然这两种病毒 DNA 的两端都具有重复序列，但这些额外序列的极性相反，所以复制期间就不能产生多基因组单位的 DNA 分子。当复制从某一端开始时，两条互补链不能同时用作模板，所以一次只能合成一条单链。合成过程也是按 5′→3′方向进行，当子链合成完成之后，其中一条亲链被替换下来，被替换下来的亲链再用作合成新子链的模板。复制也可从两端开始，从而产生复制中间体（图 6－15）。

根据对腺病毒 DNA 复制的研究，增长链的 5′端都具有一种与其末端胞嘧啶核苷酸结合的蛋白质，叫做末端蛋白（tereminal protein）。末端蛋白的分子量为 55kd，由病毒基因编码。DNA 复制之前，末端蛋白与脱氧胞苷一磷酸的 5′磷酸共价结合，然后这种复合体中的碱基 C 与模板链 3′碱基 G 进行配对，作为 DNA 合成的引物，最后由病毒基因编码的 DNA 聚合酶再合成一条连续的互补单链。

现在知道，复制终止过程需要宿主细胞的拓朴异构酶 I，以除掉因替换合成而形成的扭转的单链。如果没有拓朴异构酶 I，即使线状 DNA 的两端都已分开，也不能进行 DNA 复制。一般认为，由腺病毒基因编码的 72kd 的单链 DNA 结合蛋白在这个过程中起着类似螺旋酶的功能。

（三）痘病毒基因组的复制

痘病毒则以另一种方式复制其基因组，即复制是从基因组内部的复制启始区开始，最后通过双向复制，将原基因组两端的单链发夹环转变成双螺旋（图 6－16）。有关其复制的详细过程以及两端单链环的形成过程目前都还不清楚。可能在DNA合成完成之后，某种内切核酸酶交错切割单链环所在区域的

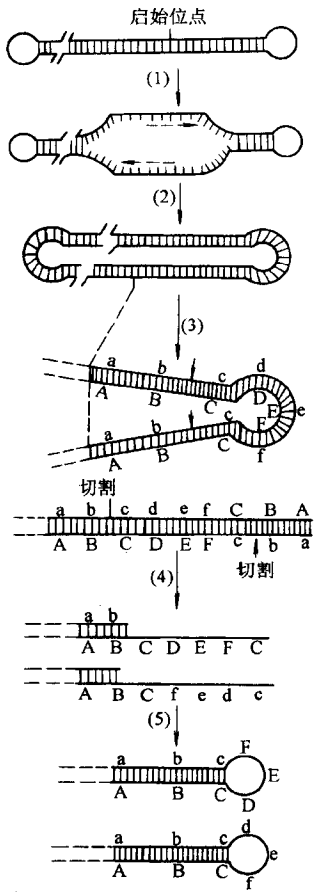


图 6－16　痘病毒 DNA 的复制

其两端具有单链发夹环结构（1）从复制启始位点开始进行双向复制；（2）将单链结构转变成双螺旋；（3）在两个位点分别进行交错切割；（4）形成单链末端；（5）单链末端序列连接成发夹环。从（3）到（5）只示基因组 DNA 的一部分

DNA，然后通过变性，使两条单链分开，最后单链尾巴再弯曲成环，由 DNA 连接酶将其连接成闭合的发夹环。痘病毒同其他 DNA 病毒的另一差别是，其复制过程是在受感染细胞的细胞质中进行。由于宿主细胞质不含 DNA 聚合酶，所以病毒必须首先合成自己的 DNA 聚合酶，然后才能进行复制。

二、环状双链 DNA 病毒基因组的复制

（一）SV_{40}基因组的复制

SV_{40}环状双链 DNA 基因组的复制在许多方面都类似细胞染色体和质粒 DNA 的复制，复制首先从惟一的一个复制起点开始，两个复制叉朝复制起点相对的一面双向进行。其中有一条链的合成按 5′→3′方向连续合成，而另一条链的合成则是不连续的，即首先合成引物 RNA，然后再合成冈崎片段，最后再由冈崎片段连接成连续的 DNA。当两个复制叉在复制起点相对的一面会合后，复制过程终止。复制终止位点的序列对完成复制过程并不重要，但复制启始信号则含有 TTAACGTTAA 序列，这段长 10 个碱基的序列位于早期基因和晚期基因之间。

SV_{40}基因组的复制几乎完全依赖于宿主细胞的复制酶，但复制启始过程需要病毒基因编码的大 T 抗原（large T antigen）或称 T 蛋白（肿瘤特异性抗原），它是早期区中惟一的一个转录单位所编码的蛋白质。早期区的初级转录子利用细胞内的加工机器进行可变剪接，产生两种 mRNA，除掉的片段较小的mRNA编码较大的蛋白质，叫大抗原（约 90kd），除掉的片段较大的 mRNA

编码较小的蛋白质，叫小 t 抗原（约 20kd）。小 t 蛋白在病毒生命周期中的作用目前不清楚，但大 T 蛋白通常存在于细胞核之中，它除了启始病毒 DNA 合成外，还刺激宿主细胞 DNA 合成，调节自身基因和晚期基因表达，与维持细胞的转化状态有关。大 T 抗原在能与三个衔接位点进行特异结合，其中一个位点位于复制启始序列之内，另外两个位点则与复制启始序列部分重叠。

（二）乙型肝炎病毒基因组的复制

引起人体和其他动物患传染性肝炎的病毒有两组，即甲型肝炎病毒和乙型肝炎病毒（hepatitis B virus）。甲型肝炎病毒为一种 RNA 病毒，而乙型肝炎病毒则为一种具有被膜的小环状双链 DNA 病毒（图 6－17）。虽然乙型肝炎病毒的基因组 DNA 呈环状，但其两条单链都是线状分子。负链中有一个缺口，缺口两侧的序列通过与正链序列互补结合成环，其 5′端与一种蛋白质共价结合。病毒粒子中的正链也有一个缺口，其 5′端是固定的，但其 3′端则变化较大，延伸的长度不等。核壳体核心中存在 DNA 聚合酶，其作用是延长不完整正链的 3′末端。

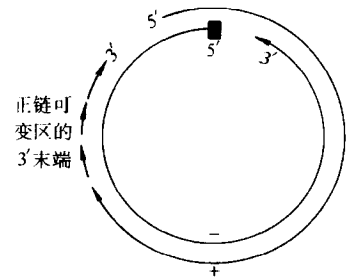


图 6－17　乙型肝炎病毒基因组（3200bp）
基因组 DNA 与 DNA 聚合酶一起被包裹到病毒粒子中，进一步延长不完整正链的 3′端，黑块为与负链 5′端共价结合的蛋白质

受乙型肝炎病毒感染的细胞含有大量与正链 DNA 结合的负链 DNA，其中大多数都比基因组 DNA 链短。负链 DNA 是在未成熟的类核心颗粒中以 RNA 为模板合成的，随着负链不断延长，RNA 模板不断降解（图 6－18）。所以乙型肝炎病毒也同逆转录病毒一样，能通过逆转录进行复制。用于负链 DNA 合成的模板 RNA 称为前基因组（pregenome），但目前对复制循环的详细过程还不清楚。同逆转录病毒不同的是，负链的启始合成可能与其 5′端共价结合的蛋白质有关，这种情况与腺病毒 DNA 的复制有些相似。正链 DNA 可能是以游离的负链 DNA 为模板合成的，其引物可能是正链 RNA 片段。

（三）植物 CaMV 病毒基因组的复制

植物 CaMV 病毒基因组的复制也与动物逆转录病毒的 DNA 合成过程有些相似，但其确切的复制机制目前还不清楚。一般认为，CaMV 基因组复制时，首先是宿主细胞的 RNA 聚合酶Ⅱ以整个病毒基因组为模板，转录成35S RNA，然后在病毒基因编码的逆转录酶的作用下，以 35S RNA 为模板合成负链 DNA。负链合成过程中的引物为宿主细胞的 tRNA，这种 tRNA 通过与

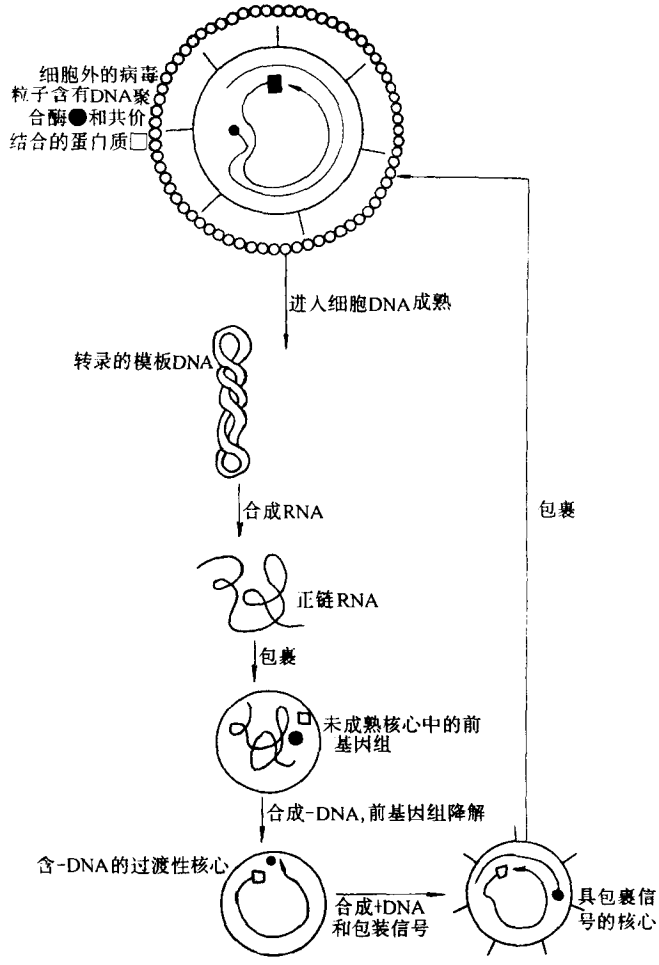


图 6－18　乙型肝炎病毒 DNA 的复制模型

35S RNA中 14 个碱基互补配对，启始逆转录过程。tRNA 的结合位点距 35S RNA 5′端约 600 个碱基，位于断裂区Ⅱ附近。当逆转录进行到 35S RNA 5′端顺式重复时，形成一种小分子负链 DNA 片段叫强终止 DNA（stong－stop DNA），长度为 600 个碱基，其 5′端仍然与 tRNA 共价结合。由于 35S RNA 具有 2 段顺式重复，所以强终止 DNA 首先使模板 RNA 形成环状分子，然后再进行逆转录，形成完整的负链 DNA。接着，复制机器再以负链 DNA 为模板，以能够与病毒 DNA 上 G2 和 G3 断裂区附近的、富含碱基 G 的序列进行互补的 tRNA 为引物合成正链 DNA。除掉引物后便形成断裂区，5′端序列再与引物位点结合形成三链结构。

三、单链 DNA 病毒基因组的复制

（一）细菌单链 DNA 病毒基因组的复制

细菌的某些小病毒如噬菌体 φX174、G4 和 M13 等需要引物 RNA 将其环状单正链 DNA 转换成双螺旋复制型（RF）。当具有感染力的正链进入宿主细胞后，由单链 DNA 结合蛋白将其包裹起来，并立即进行复制，合成一些用于启始负链合成的引物 RNA。引物 RNA 是以正链上某段特殊 DNA 序列为模板合成的（图 6－19）。已知 φX174 在复制时能够组装成完整的复制启始复合体，对 G4 噬菌体来说，引物 RNA 合成可能只需要引物酶而不需 RNA 聚合酶。噬菌体 M13 复制时，大肠杆菌的 RNA 聚合酶能够识别 M13 DNA 上小发夹环，并利用这一位点合成引物 RNA。所有这些病毒都利用宿主细胞的 DNA 聚合酶Ⅲ合成负链 DNA，然后再除掉引物 RNA，留下的缺口最后由 DNA 聚合酶 I 进行修补。

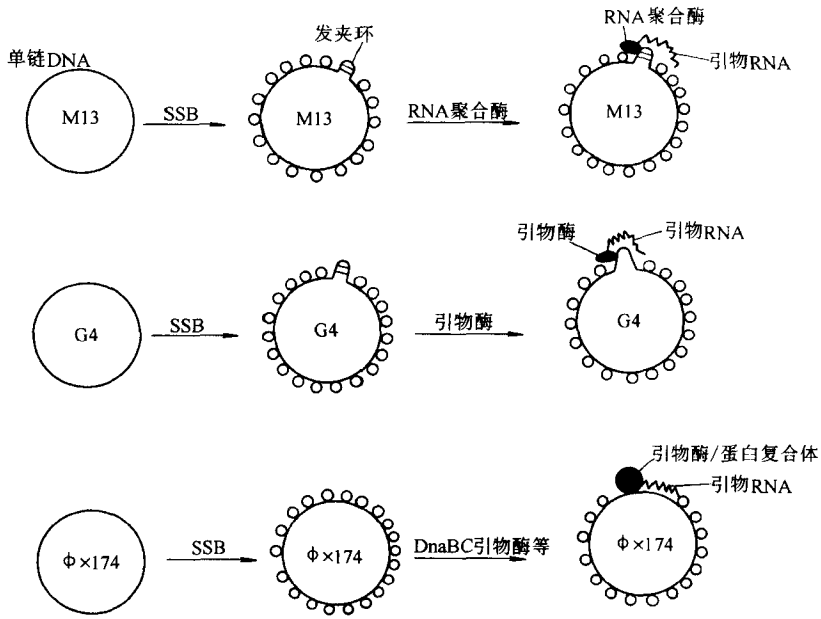


图 6－19　病毒单链 DNA 分子中合成三种引物 RNA 的机制

所有单链 DNA 噬菌体都采用滚环模型将其复制型基因组分解成能包裹到病毒粒子中去的正链 DNA，或用于进一步复制。根据对 φX174 的研究，其基因组编码一种切割酶即 A 蛋白，这种酶是启始滚环切割所必需的蛋白质。A

蛋白识别共价闭合复制型 DNA 上唯一的启始位点，并通过切割该位点与其中一条单链的 5′端共价结合。当复制叉沿正链前进时，被替换的正链的 5′端并不释放下来，而是通过切割酶与 DNA 聚合酶Ⅲ结合，作用于复制叉（图 6－20）。复制叉沿滚环运动一周后，将与其结合的切割酶带到复制起点上，进行第二次特殊切割。在进行切割和新一轮复制合成时，切割酶脱离原来正链的 5′端，并使其 5′端与 3′端连接在一起，形成一条环状的、长度为一个基因组单位的子代正链。这种切割和连接反应不需要 ATP，所以切割酶具有类似拓扑异构酶 I 的功能。

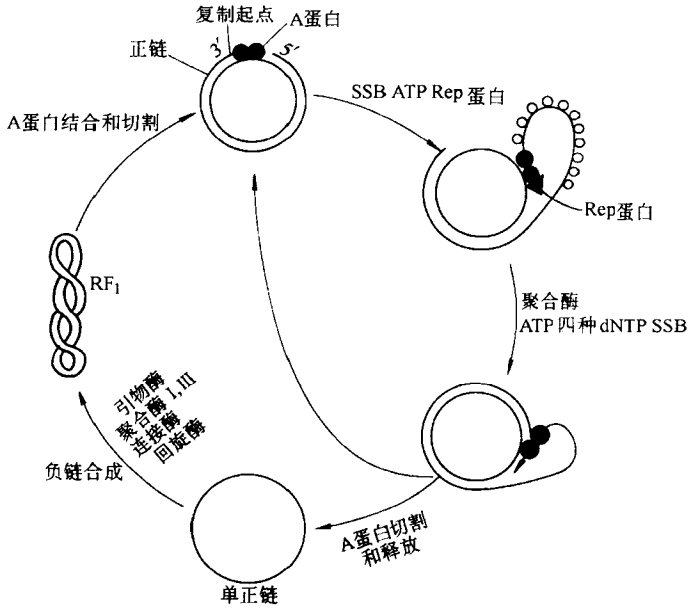


图 6－20　噬菌全 φX174 复制的两种复制型（RF）

RFI 为双链闭合环状，是以正链为模板合成互补的负链而形成的。RFⅡ为双链开环，是滚环的中间产物，用以合成新的正链

在噬菌体侵染初期合成的子代正链都通过引物 RNA 而转变成子复制型 DNA。而在噬菌体侵染后期，子代环状正链都由噬菌体特异性蛋白包裹，从而阻止引物 RNA 合成。

（二）动物细小病毒基因组的复制

细小病毒（parvovirus）为动物的一种很小的单链 DNA 病毒，其基因组长 4 500～5 000 个碱基。病毒粒子含有三种壳体蛋白，它们是由初级转录子通过可变剪接所产生的不同 mRNA 编码的。有一组细小病毒只能在受其他一些病毒如腺病毒感染的细胞中生长，其余细小病毒具有自主性，其生长不依赖于其

他病毒的感染，但只有在进行细胞分裂迅速的细胞中才能复制。在自然条件下，复制自主性病毒只能感染某些类型的细胞，如迅速增殖的成血细胞。这些病毒也能感染肿瘤细胞，因为肿瘤含有迅速增殖的细胞。

细小病毒基因组的末端序列非常特殊，能够折叠成一种特殊的T型末端结构（图6－21），以便于DNA分子末端的复制。基因组的3′末端序列可用作引物，当加进4种脱氧核糖核苷酸后就可产生一个拷贝的基因组，如果末端再次折叠又形成下轮复制的引物。当单链在某些特殊位点形成缺刻后，于是产生新的基因组DNA。

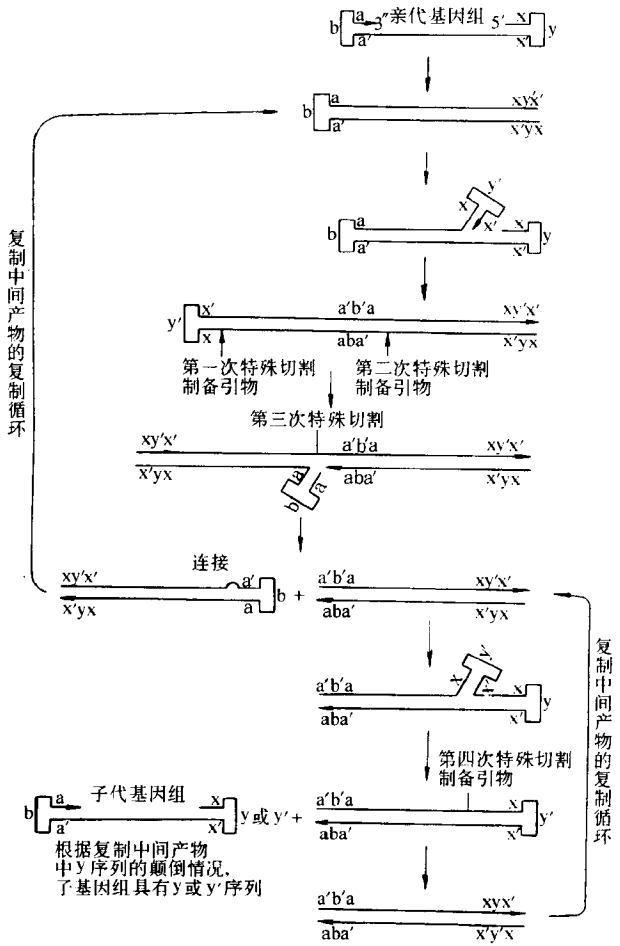


图6－21　细小病毒的复制模型

复制过程中交替地以两种发夹环为引物

四、单链 RNA 病毒基因组的复制

大多数单正链 RNA 病毒基因组的复制都需要一种 RNA 复制酶（RNA replicase），这种酶是在病毒 RNA 进入宿主细胞后合成的。同 DNA 聚合酶和 RNA 聚合酶相似，RNA 复制酶催化单链模板 RNA 链的互补合成，其合成机制与 DNA 复制和 RNA 转录基本相同。

当双链 RNA 形成之后，互补的两条单链又作用下一轮 RNA 合成的模板。合成过程中有一部分子链被包裹到新的病毒粒子之中，另一部分则用于指导合成某些特殊蛋白质。由于两条互补链中只有一条单链用作某个基因的 mRNA，所以 RNA 和 DNA 病毒 mRNA 的形成都有一定相似性。

大多数单链 RNA 病毒在繁殖期间，编码蛋白质的链均与病毒粒子中的 RNA 链相同，所以具有侵染力的 RNA 链编码 RNA 复制酶。但是动物和植物的某些单链 RNA 病毒只有一条负链 RNA，所以具有感染力的 RNA 不能编码复制酶。根据对动物 RNA 病毒水泡性口炎病毒的研究，虽然具有感染力的基因组 RNA 不能编码复制酶，但每一 VSV 病毒粒子中都包裹了一条复制酶分子（L－NS 蛋白复合体），所以当负链基因组 RNA 进入宿主细胞后就能立即合成大量 mRNA。这条复制酶分子是在前一轮病毒侵染循环中合成的，它随病毒 RNA 一起进入宿主细胞并立即将负链复制成用作 mRNA 的正链。

五、逆转录病毒基因组的复制

（一）逆转录病毒 DNA 的合成

虽然逆转录病毒的基因组为正链 RNA，但复制的中间产物都是基因组的一种双链 DNA 拷贝。双链 DNA 整合到宿主细胞的染色体上后就类似宿主细胞的基因，通过转录，形成新的病毒基因组。所以新病毒基因组是在宿主细胞核中、由细胞内有关机制合成和进行加工的。虽然病毒基因组 RNA 与 mRNA 相似，但在病毒感染的初期，其基因组并不行使 mRNA 的功能，只是用作合成 DNA 的模板。

逆转录病毒粒子中的逆转录酶依据单链 RNA 复制成的双链 DNA 分子在两端都接有叫做长末端重复（LTR）的序列。LTR 由 U3－R－U5 组成，即 R 序列两侧衔接有一段 3′端序列和一段 5′端序列，其中含有 RNA 聚合酶Ⅱ的转录启始信号以及转录子 3′端的切割信号和多聚 A 化信号。

病毒粒子在进入宿主细胞后不久就在细胞质中进行 DNA 合成。首先是逆转

录酶从 tRNA 引物开始加上脱氧核糖核苷酸，形成一段长 100～200 个碱基的 DNA 序列，即 R 区和 tRNA 引物之间与 R 区和 U5 互补的一段序列（图 6－22）。然后逆转录酶将杂种分子中的 RNA 除掉，使新合成的 DNA 片段与基因组 RNA3′端的 R 序列进行碱基互补配对，形成一段新的模板/引物结构。因此，逆转录酶就能继续复制基因组 RNA 的其余序列。正链合成之前，必须在 RNA 中 U3 的左侧切割，产生具有 3′－OH 末端的 RNA，然后逆转录酶再以此作为引物合成一段正链 DNA。当逆转录酶降解与负链 DNA 相连的 tRNA 和基因组 RNA 的其余部分后，正链片段中相当于引物结合位点的序列与负链 DNA 的相应部分进行碱基互补配对，完成两条单链的合成，最后产生线状双链结构。

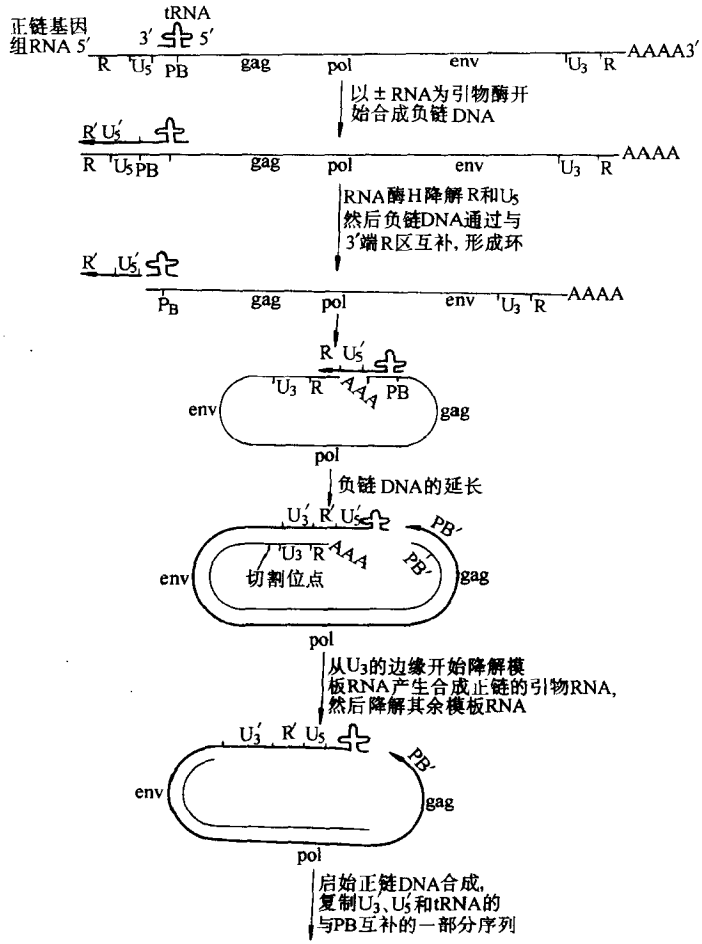


图 6－22（1）　逆转录病毒 DNA 的合成（一）

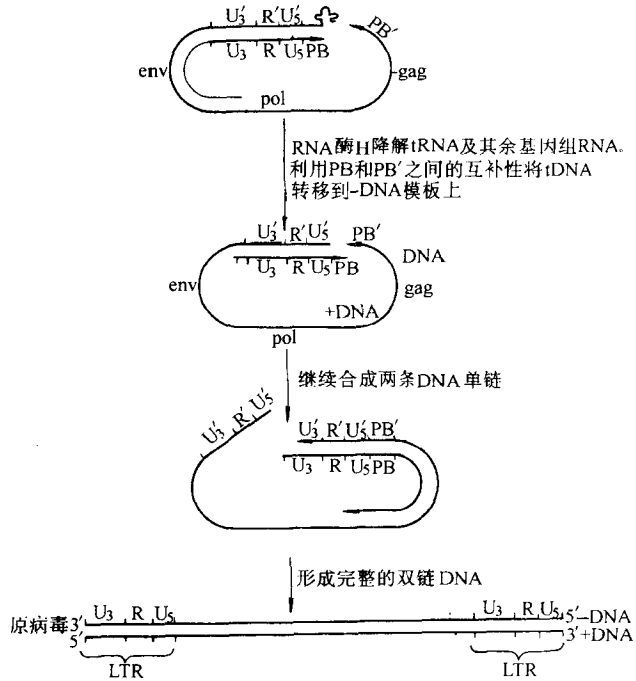


图 6－22（2）　逆转录病毒 DNA 的合成（二）

所有步骤均由逆转录酶催化

原病毒 DNA 合成的各个步骤都是由逆转录酶催化的，所以逆转酶至少具有 4 种酶活，即依赖于 RNA 的 DNA 聚合酶活性、RNA 酶 H 活性、依赖于 DNA 的 DNA 聚合酶活性和在 U3 的 5′端切割 RNA 的特殊的内切核酸酶活性。线状 DNA 合成之后被运输到细胞核中，宿主细胞的某种酶将其转换成共价结合的闭合环状分子，然后再整合到宿主染色体上。病毒 DNA 的整合机制目前还不很清楚，但现在知道，整合期间环状原病毒必须打开成线状分子，因为整合到染色体上的病毒 DNA 的两端始终相同。相反，染色体上的整合位点则无位点特异性。在病毒 DNA 与宿主 DNA 之间的连接处，原病毒的两端都发生 2 个碱基的缺失，而整合位点上的宿主 DNA 则有 6 个碱基的重复（图 6－23）。整合过程至少需要一种由病毒基因编码的内切核酸酶，即所谓整合酶（integrase）。

整合到宿主细胞 DNA 上的原病毒很稳定，并随宿主 DNA 的复制而进行复制。这种特征同 λ 噬体有些相似，但不同的是逆转录病毒的原病毒不会从整合位点上切割下来。逆转录病毒感染细胞后通常不会损害细胞，受感染的细胞继续分裂，而且原病毒 DNA 通过转录合成病毒 RNA。爱滋病毒感染则属例外情况。爱

滋病毒能够感染某些特殊的帮助 T 淋巴细胞（helper T lymphocyte）并将这些细胞杀死，所以爱滋病患者由于缺少帮助 T 细胞而丧失某些免疫功能。

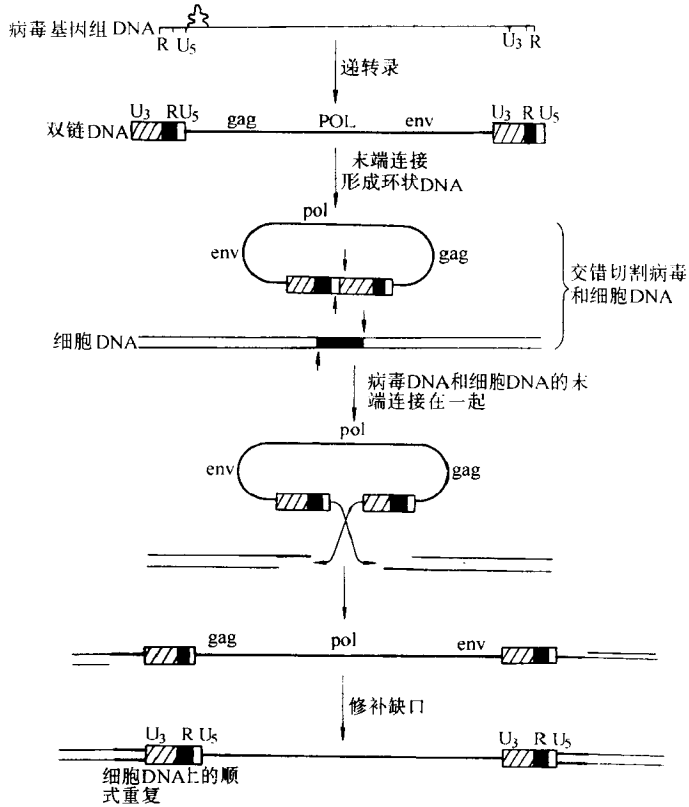


图 6－23　逆转录病毒 DNA 的整合模型

病毒基因编码的内切核酸识别由颠倒重复连接在一起形成的序列，整合期间丢失最后两对碱基

（二）原病毒通过转录形成病毒粒子

长末端重复不仅提供整合功能，而且还含有转录信号。LTR 的 U3 序列含有控制转录的启始信号，在 U3－R 相邻处的帽子位点上游－25 和－80 的位点上分别具有典型的 TATA 框和 CCAAT 框，在更上游的位点上还存在长 70～100bp 的增强子顺式重复序列（图 6－24）。逆转录病毒的增强子与 SV_{40} 的增强子非常相似，在某些情况下还可相互替换。整合到宿主染色体上的原病毒由 RNA 聚合酶Ⅱ转录，转录子由细胞内的加工机器进行加工。转录子在 R 区的 3′端被切割后再接上多聚 A 尾巴，成为成熟的 mRNA。其中一部分被运到细胞质中，组装成病毒粒子，成为新的病毒基因组。另一部分则被运到细胞质中作为 gag 和 pol 基因的 mRNA（图 6－25）。还有一部分则通过剪接形成 env 基

因或 px/lor 基因的 mRNA。

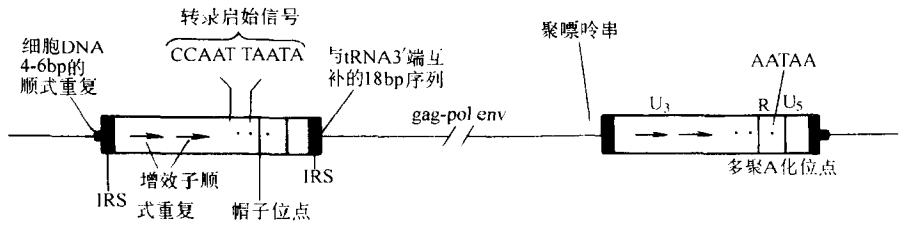


图 6－24　LTR 的特征

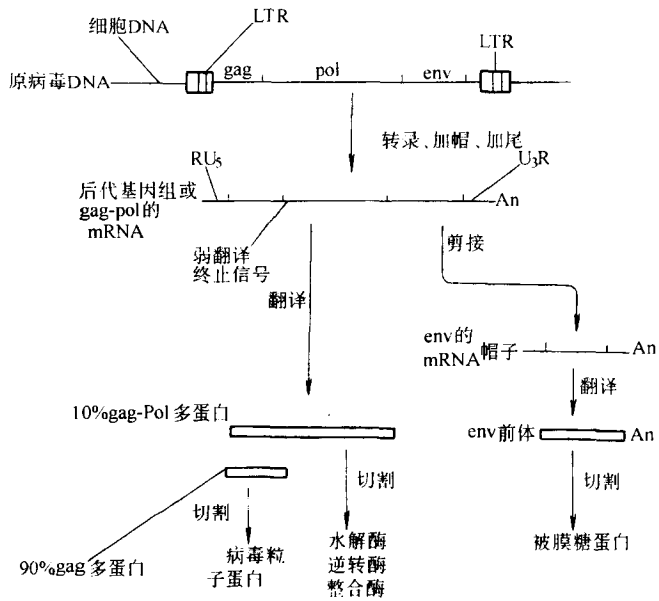


图 6－25　逆转录病毒 RNA 和蛋白质的合成

一部分初级转录子通过剪接形成 env 的 mRNA，其余则用作基因组或 gag mRNA，只有一小部分用作 gag－pol 的 mRNA，即通过 gag 末端的终止密码抑制下游顺反子翻译

（三）生殖系中的原病毒

根据前述复制循环，逆转录病毒不仅可以成为感染因子，而且还可以成为宿主细胞的基因。如果病毒感染了某个生殖细胞，那么原病毒就可传递到子代，作为细胞内一种正常基因进行遗传，这种遗传因子叫做内源原病毒（endogenous provirus）。在某些动物如小白鼠中，其基因组约含有数千条内源原病毒，其 DNA 含量约占总体 DNA 含量的 0.5%。事实上，在许多脊椎动物包括人体 DNA 中都存在这种内源原病毒。

内源原病毒在结构上与外源逆转录病毒感染的原病毒很相似，如果内源原病毒通过各种化学制剂的诱导，可变成具有感染力的病毒。某些化合物如5－溴脱氧尿苷、碘代脱氧尿苷（iododeoxyuridine，IUdR）的诱导效率非常高。内源原病毒通常不转录，其部分原因是其GC序列都被甲基化，由于诱导因子能够逆转甲基化作用，所以能够使内源原病毒变成具感染力的病毒。

同外源逆转录病毒一样，内源原病毒在染色体DNA上的分布也是随机的。原病毒插入到某个基因中可以使该基因失活，使个体具有一定的表型特征。例如小鼠的d座位中插入了原病毒后，其皮毛呈淡褐色。当然，D座位也可发生回复突变，这种由原病毒引起的回复突变频率在每个世代中约为3×10^{-6}。

现在我们能够将新的原病毒有目的地插入到实验小白鼠的生殖细胞中，即通过在离体条件下感染小鼠的早期胚胎，然后再将其移植到寄养母体内，使其进行正常发育。以这种方式培育的动物在其许多不同类型细胞的不同染色体位点上都载有原病毒。通过对这些小白鼠进一步选育，可以培育出原病毒按孟德尔方式遗传的新品系。这种原病毒非常类似内源原病毒，新引入的原病毒有时会使插入位点的基因失活，从而引起突变。根据以上所述，内源原病毒在许多特征方面类似转座子。真核细胞中有许多转座子都与原病毒相似，如果蝇的Copia家族，酵母的Ty因子等，所以这些转座子都称为逆转座子（retrotrasposon）。事实上，原病毒和逆转座子的复制和插入机制都与逆转录病毒的机制相同。

第三节　病毒基因的转录及蛋白质合成

一、病毒基因的转录启始信号

（一）DNA噬菌体基因的启动子

1. T噬菌体基因的启动子与转录　双链DNA噬菌体基因组的表达都是分阶段进行的，也就是说噬菌体基因组在复制循环过程中分阶段地进行转录。对于T系列噬菌体而言，当其基因组DNA进入宿主大肠杆菌细胞以后，宿主细胞中的RNA聚合酶将一小段病毒DNA转录成mRNA，开始合成早期蛋白，所以宿主细胞的RNA聚合酶能够识别早期蛋白质基因的启动子。但是事实上，所有T噬菌体早期基因的启动子的结构都与其宿主大肠杆菌基因的启动子有很大差别。虽然早、晚期基因在转录起点上游都具有相当于细胞基因－10框的序列，但相似程度很低，而且各个基因的启动子之间都存在比较明显的差别。例如在T7噬菌体中，虽然有2个早期基因（A1和A0）启动子非常相似（表6－1），都具有同感序列AT－CT－A，但这段序列均与大肠杆菌基因的启

动子有很大差别。另外2个早期基因的启动子A2和A3以及T4噬菌体的基因63的启动子则与大肠杆菌“－10”位的TATAAAT框有某种程度的相似性。比较特殊的是，在T7的所有早期基因启动子的上游都存在一个翻译终止信号TAA，目前尚不清楚这种终止密码是否对转录启始过程具有作用。

表6－1　T噬菌体早期基因的启动子

	基　因				启动子	起点
T7	A0	ACCTCC	TAA	CGTCC－－－	ATCCTAA	AGCCAAC＊ACC
	A1	AAAGTC	TAA	CCTATAGG	ATACTTA	CAGCCAI＊CGA
	A2	ATGAAG	TAA	CATGCAG－	TAAGATA	CAAATCG＊CTA
	A3	ATGAAG	TAA	ACACGG－－	TACGATG	TACCACA＊TGA
T4						
	基因63	TCCCTC	GTG	TTGTGT	TATAGTA	GTCTTAC＊TGA

根据表6－1知道，T7噬菌体的每一个早期基因的转录起点的碱基各不相同，所以早期基因实际的转录起点变化很大，但所有基因的转录起点都接近翻译起点密码ATG。

T噬菌体的中、晚基因都由噬菌体基因编码的RNA聚合酶转录，这种酶只有一条多肽链，比宿主细胞的RNA聚合酶简单。但是，噬菌体基因转录除了需要这种RNA聚合酶以外，还需要其他一些转录辅助因子，如T4噬菌体中晚期基因转录至少还需要另外4～5个基因的编码产物。由于这些转录辅助因子的作用很复杂，对其特性，目前知道得不多。

在各种T噬菌体中，T7的中、晚期基因的启动子都不具有大肠杆菌基因的启动子的典型特征（表6－2），其共有序列为TAATACGACTCAC-TATAGGGAGA。这些基因都由早期基因之一编码的RNA聚合酶转录，这种酶非常活跃，对中、晚期基因的启动子具有高度特异性。但是在T4的三个中期基因中，转录起点上游都具有TAT－AAT序列，这种结构比较接近大肠杆菌基因“－10”框的同感序列。在噬菌体T3中，晚期基因的启动子区含有结构保守序列T－ACCCT，某些晚期基因在这段序列上游有三个碱基TAA，如果将这两部分合起来就构成TAAT－ACCCT，显然这种序列非常类似大肠杆菌基因“－10”框的序列。T3晚期基因在这一位点下游的序列的保守性程度也比较高，具有CACATT序列，但这段序列与其他噬菌体中、晚期基因相同位置的序列无任何相似性。

表 6－2　T 系列噬菌体中晚期基因的启动子

基　因					启动子 −10		起点	
T7	中							
	Φ1.5	AGTTAA	CTC	GTAATA	CGACTCA	CTAAAG	G	AGG
	Φ1.6	TGGTCA	CGC	TTAATA	CGACTCA	CTAAAG	G	AGA
	Φ2.5	GCACCG	AAG	TAA－TA	CGACTCA	CTATT－	A	GGG
	Φ3.8	CGTGGA	TAA	TTAATT	GAACTCA	CTAAAG	G	GAG
	Φ4c	CCGACT	GAG	ACAATC	CGACTCA	CTAAAG	A	GAG
	Φ4.3	AGTCCC	ATT	CTAATA	CGACTCA	CTAAAG	G	AGA
	Φ4.7	TTCATG	AAT	ACTATT	CGACTCA	CTATAG	G	AGA
T4	中							
	denV	TACATC	TCC	TGTAGG	TATGATA	CTATAG	A	CCT
	基因 55		（不完整）		TATGAAT	TGAGCT	A	AGA
	orf D	GCTCCT	ATA	TTGCTT	TATAAAT	TTTT－－	T	GGT
T7	晚							
	Φ6.5	GTCCCT	AAA	TTAATA	CGACTCA	CTATAGG	G	AGA
	Φ9	GCCGGG	AAT	TTAATA	CGACTCA	CTATAGG	G	AGA
	Φ10	ACTTCG	AAA	TTAATA	CGACTCA	CTATAGG	G	AGA
	Φ13	GGCTCG	AAA	TTAATA	CGACTCA	CTATAGG	G	AGA
	Φ17	GCGTAG	GAA	ATAATA	CGACTCA	CTATAGG	G	AGA
	ΦOR	CACGAT	AAA	TTAATA	CGACTCA	CTATAGG	G	AGA
T3	晚							
	PjB10	AAACAC	TGG	AAGTAA	TAACCCT	CACTAA	C	AGG
	HpaIN	TCCAAC	GTT	GTCTAT	TTACCCT	CACTAA	A	GGG
	PjB20	GAAGTG	AAA	GCCTAA	TTACCCT	CACTAA	A	GGG
	NbOI－E	TCAATG	AGT	TTGCAT	TAACCCT	CACTAA	A	GGG
T4	晚							
	基因 67	TCGTTT	CCA	AGACCC	CGACCAA	GAACAA	G	AGG
	Orj		（不完整）		－AAGCTT	GCTAAG	G	AGA
	P23	CACTAT	TAC	TGAGAG	TATAAATA	CTCCCT	G	ATA
	基因 45	TTTAAC	+15	AAATTA	GTTATAA	AATTAA	A	TCT

注：与 −10 框相邻的序列也可能对转录启始具有作用。

2．*其他双链DNA噬菌体基因的启动子与转录*　同前述T噬菌体基因的启动一样，λ噬菌体基因的启动子也与其宿主大肠杆菌基因的启动子不存在相似性或同源性（表6－3）。根据对λ噬菌体5个研究较清楚的启动子序列的比较分析，这种差异在“－10”区表现更为明显，而且这5个启动子之间也无相似性。但是在“－35”位，启动子PI和P_{RE}之间以及P_R和P_{RH}之间存在较高的同源性，而其他启动子之间则不存在这种同源性。此外，由于转录起点的碱基均为A或T，可能RNA聚合酶倾向性地选择这两种碱基作为转录起点。由于这些基因的控制表达过程很复杂，需要几个顺反子编码的产物，特别是与溶原性有关的顺反子更是如此。有关这些启动子的功能将在第八章中加以讨论。

表6－3　部分λ噬菌体基因的启动子

	（－35）		（－10）		起点
大肠杆菌共有序列	TGTT－GACANTTT		TATAATG		
P_R	GTGTTGACTATTTTA	CCTCTGGCG－－	GTGATAA	TGGTTGC	A　TGT
PI	TTGC－GTGTAATTGC	GGAGACTTTGC	GATGTAC	T－－－－－	T　G－－
P_{RE}	TTGC－GTTTGTTTGC	（不完整）	GTAAGTA	T－－－－－	A　G－－
PaQ	GCTC－GTGAACGTCA	TGGAAACGGA－	ATCATAA	AGGAAGT	T　CGA
P_{RM}	GTGTTAGATATTTAT	CCCTTGCGGTG	ATAGATT	TAA－CGT	A　TG－

枯草杆菌（*Bacillus subtilis*）的一种噬菌体SPO1基因组转录时，需要用5－羟甲基尿嘧啶替代其基因组DNA中的胸腺嘧啶，这种替代过程是病毒基因转录的一个极其重要的步骤。SPOI噬菌体的基因tf1编码一种叫做转录因子TF1的DNA结合蛋白，TF1蛋白倾向性地同病毒DNA中被修饰碱基的位点互作，对转录过程进行负控制。SPO1噬菌体基因的转录需要三种不同的RNA聚合酶，宿主RNA聚合酶转录早期基因，中期基因的转录酶为噬菌体基因编码的δgP28蛋白，晚期基因的转录则与噬菌体基因编码的两种蛋白质δgp33和σgp28有关。各个基因的启动子也都与典型的原核生物基因启动子不存在同源性。

枯草杆菌的另一种噬菌体即φ29的G3b基因的启动子则与大肠杆菌基因“－10”位的序列存在较高同源性，“－35”位的序列则与SPO1噬菌体早期基因的相应序列存在较高同源性。例如G3b基因“－35”位的序列为AATTGTCGAACA，而SPO1早期基因则为AAGTTTCAAACA，二者只有4个碱基的差别。此外，这两种噬菌体基因的核糖体识别位点的结构也很相似，在SPO1的顺反子中存在AAAGGGTGG序列，而φ29噬菌体的早期基因的核

糖体结合位点则为 AGGAGG 序列。

根据对大肠杆菌的另一种噬菌体 P2 的 4 个晚期转录单位的分析，P2 晚期基因的表达需要宿主 RNA 聚合酶和噬菌体基因 ogr 编码的产物，同时也需要 P2 基因 A 和 B 编码的复制酶，其中 ogr 基因产物修饰宿主 RNA 聚合酶的 α－亚基。P2 晚期基因的启动子序列差异很大，都不与大肠杆菌启动子存在同源性。

3．*单链 DNA 噬菌体基因的启动子*（virus gene promotor）　有关单链 DNA 噬菌体基因的启动子目前知道得不多。根据对 φX174 和 fd 噬菌体的分析，基因组中有一个转录识别信号 P（x）与大肠杆菌基因的启动子的共有序列非常相似，其余基因的启动子的同源程度都很低。

（二）真核生物 DNA 病毒的启动子

1．*腺病毒基因的启动子*　腺病毒所有早期基因都由宿主细胞的 RNA 聚合酶Ⅱ转录，其中 EIA 基因可能是最早表达的一个早期基因，因为它是与其他早期基因转录有关的一个关键基因。EIA 基因在转录起点上游－28 位具有 TA－TA 框，这种序列同宿主基因的启动子相似（表 6－4）。虽然在这一区域的上游不存在结构保守序列，但这一区域缺失后会严重降低 EIA 基因的转录水平，因此它可能对转录起着某种程度的调节作用。EIA 基因编码长 289 个氨基酸的磷蛋白，是其它任何早期基因转录所必需的。

表 6－4　哺乳动物 DNA 病毒基因的启动子

	辅助信号			启动子			起点	
腺病毒Ⅱ								
E1A	CGTTTTTATT	ATTTATAGTCAGC	TGACGTGTAGTG	--TATTTATA-	CTCGGTGAG	(+10) GCC	A	CTC
E2A1	AGATGACGT	AGTTT (+23) CG	CGAAACTAGTCC	--TTAAGAGT-	CAGCGCGCA	(+8) TGA	A	GAG
E2A2		(不完整)	TCAGG	--TACAAATT-	TGCGAAGGT	(+9) TCC	A	CAG
E2A3	AGATGACGT	AGTTT (+21) AG	GGCGCGAAACTA	--GTCCTTAA-	GAGTCAGCG	(+12) TGA	A	GAG
E2A3	CGCCGGGTG	TGGCC (+6) ACG	TAGTTTTCGCGC	--TTAAATTT-	GAGAAAGGG	(+13) CTT	A	AGA
E3	GGCGCAGCT	TGCGG (+23) TC	GCCCGGGCAGGG	--TATAACTC-	ACCTGAAAA	(+9) GAG	G	TAT
ML (晚)	GTGATTGGT	TTATA (+24) TC	CTGAAGGGGGGC	--TATAAAAG-	GGGGTGGGG	(+11) CTC	A	CTC
IVa2 (中)	CCCCCTAGT	GGACA (+15) CC	CACTTAGCCTCC	--TTCGTGCT-	GGCCTGGAC	(+11) GTC	T	CAG
SV40								
φ早期	CCATTCTCC	GCCCCATGGCTG	ACTAATTTTTTT	--TATTTAT--	GCAGAGGCC	GA (+7) TCG	G	CCT
φ晚期	(不完整)	CAGCTGGTTCTT	TCCGCCTCAGAA	GGTACCTAACC	AAGTTCCTC	(+8) GTT	A	T

早期基因 E2A 表达需要 E1A 蛋白的调节，E1A 基因有 2 个启动子，控制两种 mRNA 的转录，其中较小的一条 mRNA（12S）的编码产物抑制 E2A 启动子的活性。E2A 启动子同 E1A 启动子有部分同源性，但二者至转录起点的

距离略有不同。E2A 区有 3 个启动子即 E2A1、E2A2 、E2A3，三者结构各不相同。后来发现 E2A3 也包含 2 个启动子，即 E2A3Φ1 和 E2A3Φ2，后一个启动子距转录起点较远，与标准 TATA 框存在较高相似性，而前一个则与标准启动子不存在同源性。其他早期基因的启动子之间也都存在很明显的差别。尽管腺病毒各个早期基因的启动子的碱基序列存在很大差别，但所有这些基因都由宿主细胞的 RNA 聚合酶Ⅱ转录，可能这种转录酶对病毒启动子的识别过程需要其他辅助蛋白的帮助。

腺病毒的中、晚期基因由宿主细胞的 RNA 聚合酶Ⅲ转录，但也有一些编码 mRNA 的晚期基因同早期基因一样由 RNA 聚合酶Ⅱ转录。对这些中、晚期基因的转录机制目前还不很清楚，但现在对其中一个主要的晚期基因（ML）和一个与其相连的中期基因（IVa2）的启动子了解比较清楚。这两个顺反子的转录方向相反，ML 基因在转录起点上游有一个 TA－ TA－框，而中期基因则不具有这种结构。

2. SV_{40}基因的启动子及其转录　SV_{40}的早期基因在病毒的侵染循环过程中始终如一地表达，而晚期基因只是在病毒开始复制后才表达。当病毒侵入宿主细胞核或完全脱去蛋白质外壳后，宿主细胞的 RNA 聚合酶Ⅱ从早期基因的惟一的一个启动子开始转录，产生 2 种编码大、小 T 抗原的 mRNA。当这些 mRNA 被翻译后，大 T 抗原回到细胞核内，特异性地结合到病毒基因组上，与自身基因的启动子、晚期基因的启动子和 DNA 复制启始区互作，启始病毒 DNA 复制和控制早、晚期基因的转录。如前所述，大 T 抗原还诱导宿主基因组中某些复制酶基因表达，从而诱导宿主 DNA 的复制。此外，还有一小部分大 T 抗原穿过宿主质膜，行使一种肿瘤特异性转移抗原的功能。

SV_{40}早期基因的启动子序列同腺病毒 E1A 基因的启动子相似，但有一个重要区别，即 SV_{40}早期启动子上游有 7 个连续的碱基 T。启动子与转录起点相距 9bp 时，转录效果最佳，因此这一特性与原核生物基因的启动子相似。

除了启动子以外，增强子对早、晚期基因的转录也都具有重要作用。增强子由一对相同的、长 72bp、首尾连接的重复序列组成，其中一段 GT 序列 5′ GGTGTGGTTTAAAG3′可能具有特殊作用。此外，启动子上游一段 21bp 的、富含碱基 GC 的重复序列（表 6－4）对转录也具有重要作用。总而言之，SV_{40}早期基因的转录需要宿主 RNA 聚合酶Ⅱ、启动子、增强子、大 T 抗原以及 21bp 的 GC 重复序列。SV_{40}早期基因的转录并不是一个简单的半自主过程，而是一个相当复杂的过程，需要许多参与互作的酶类和各种结构成分。

SV_{40}的晚期基因编码两种 mRNA，一种为 VP1 蛋白的 mRNA，另一种为 VP2 和 VP3 蛋白的 mRNA。这两个晚期基因也只有一个启动子，其长度为

11bp。同早期基因的转录一样，晚期基因转录也需要前述同一种增强子，但增强子以相反的方向作用于晚期启动子；同时也需要早期基因转录所需的 21bp 的 GC 序列。由于早期基因的转录控制机制也同样影响晚期基因的转录，所以晚期基因的调节表达就显得更为复杂。现在知道，至少有一个大 T 抗原亚基参与调节晚期基因的转录，还知道在晚期基因转录期间，有一种未知的因子与非模板链结合，使 DNA 的构型发生改变。

（三）某些 RNA 病毒基因的转录

在本章前述某些 RNA 病毒中，其基因组为单正链 RNA 分子，所以在基因表达过程中没有必要进行转录。但对基因组为负链 RNA 的病毒来说，在基因组的复制以及基因表达过程中，都需要将遗传信息转录成可翻译的 mRNA。

1. *水泡性口炎病毒基因组的转录*　所有负链 RNA 病毒包括水泡性口炎病毒的基因组都被包裹在核蛋白外壳之中，形成核壳体，并以这种结构作为复制的模板和转录的模板（图 6－26）。这两个过程都由同一种病毒 RNA 聚合酶

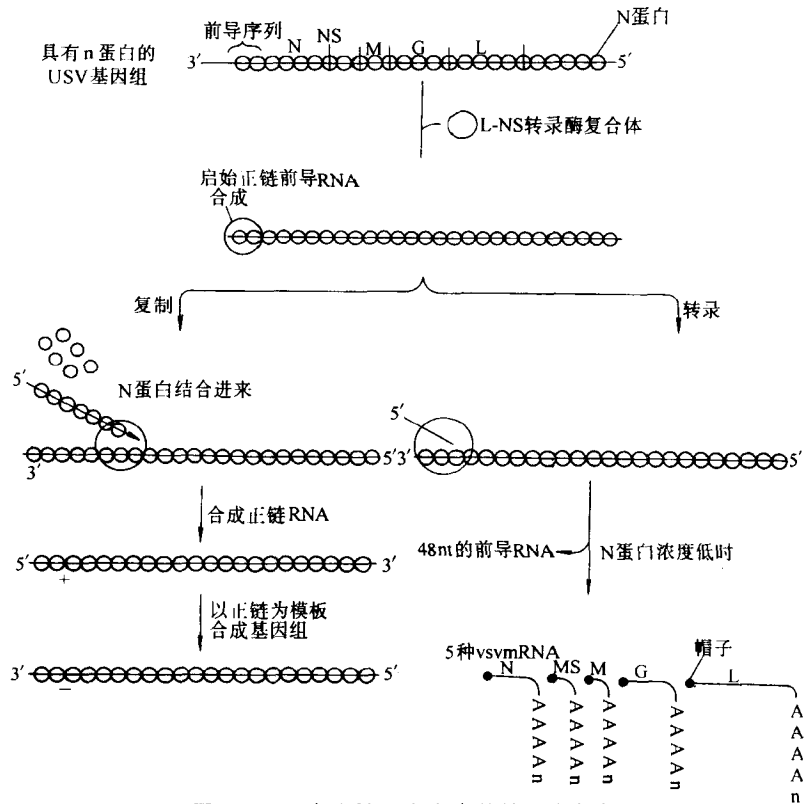


图 6－26　水泡性口炎病毒的转录和复制

进入何种途径受 N 蛋白的浓度控制

所催化，可能也需要宿主因子的帮助，但向何种途径转变，则取决于核壳体蛋白（N）的浓度。在形成病毒粒子的过程中，转录酶被包裹到病毒粒子之中，转录酶从病毒基因组的3′端开始，启始转录过程。转录是按5个基因的排列顺序连续进行的，即3′N－Ns－M－G－L5′，这5个基因共用一个功能启动子，该启动子位于基因组的3′端。转录首先从基因组的3′端开始，产生一段前导RNA。前导RNA中也有一个顺反子，其功能可能是关闭宿主细胞某些大分子化合物的合成。

水泡性口炎病毒的5种mRNA均采用一种特殊机制加上多聚A尾巴。负链基因组在相当于各条mRNA之间的区域都具有一段特殊的序列，其中存在7个碱基U。在转录过程中，转录酶在这段序列之间前后滑动，将一段多聚A序列加到mRNA3′端（图6－27）。

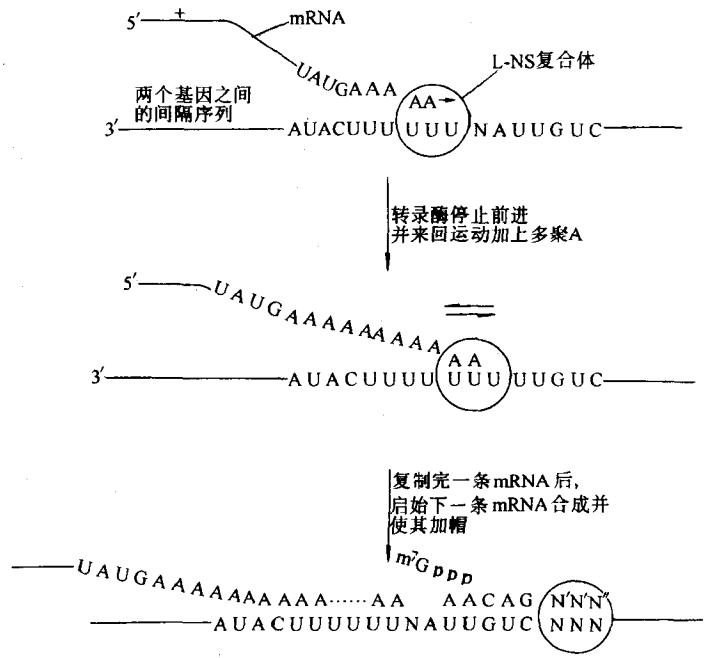


图6－27　水泡性口炎病毒mRNA的多聚A化

所示序列为两个相邻基因之间的序列

正链RNA的合成可以从转录状态转变到复制状态，即合成基因组RNA的完整的模板。这一转变过程受N蛋白调节。当N蛋白充足时，新合成的正链RNA部分立即与N蛋白结合，在转录酶之后形成正链核壳体。N蛋白与正链结合就会阻止其多聚A化，从而合成一条完整的正链RNA分子。由于负链

基因组都与N蛋白结合，转录酶只能从前导RNA区域启始RNA合成，当前导RNA片段合成结束并从模板上释放下来后，如果缺少N蛋白，于是转录酶继续合成mRNA而不能合成完整的正链RNA。根据上述模型，当N蛋白充足时，转录酶复制基因组；如果N蛋白不足，转录酶自动调整，合成更多的mRNA，从而合成更多的N蛋白。

2. *流感病毒的转录*　如前所述，流感病毒的单负链RNA基因组包括8条RNA分子，其中有3条（第6～8条）为二顺子反子结构，其余均为单顺反子。流感病毒基因组的转录需要帽子结构，需要核壳体蛋白（NP）和三种P蛋白，即两种碱性蛋白PB1和PB2以及一种酸性蛋白PA。三种P蛋白结合在一起形成复合体，即所谓的转录酶。在病毒开始转录时，这三种蛋白质作为一个整体从病毒基因组RNA链的3′端移到mRNA上。PB2也与帽子结构互作，然后病毒内切核酸酶在距帽子位点10～13个碱基的位置切割一个嘌呤碱基，形成带有帽子的片段。帽子片段是转录的引物，转录机器通过在引物3′端加一上个鸟嘌呤而启始转录过程。然后RNA聚合酶复合体按常规方式延长mRNA。因此，在转录过程中，启动子序列并不是决定因素，可能具有帽子结构的单链基因组RNA才是控制转录的主导因素。流感病毒基因组RNA本身并不具有帽子结构，帽子基团是从新合成的宿主细胞mRNA上获得的，所以当宿主细胞的mRNA合成受到抑制时，流感病毒mRNA的合成也迅速下降。

3. *植物RNA病毒基因的转录*　植物RNA病毒基因的转录以及亚基因组RNA的形成过程目前都不清楚。对于单正链RNA基因组病毒来说，如烟草花叶病毒、萝卜黄化花叶病毒等，这些病毒基因组RNA的5′端都具有帽子结构，3′端都具有多聚A尾巴。虽然这些正链RNA分子没有必要进行转录，但是其亚基因组RNA链则需要转录。如烟草是花叶病毒的编码30kd蛋白的亚基因组和17kd外壳蛋白的亚基因组都是负链RNA，所以都需要转录。前者从距亚基因组3′端1 550个碱基的鸟嘌呤核苷酸位点开始转录，后者从距亚基因组3′端693个碱基的位置开始转录。

二、病毒基因的转录终止信号

有关病毒基因的转录终止（virus transcriptional termination）过程目前还不很清楚，研究资料主要来自噬菌体。

（一）噬菌体基因的转录终止子

1. *T噬菌体*　根据对T7基因组的序列分析知道，其基因组有2个转录终止子（表6-5）。其宿主细菌的转录终止序列通常含有一串碱基T或一个双

重对称结构，而整个 T7 基因组却不存在这两种结构。但这两种高度同源的序列 TE 和 Tφ 具有终止转录的功能，其中 TE 为早期基因的转录终止子。

2. 丝状噬菌体基因的转录终止子　在三种丝状噬菌体（f1、fd 和 M13）中，转录终止子的 DNA 序列都高度同源，并且都具有一个双重对称结构（斜体）。双重对称结构都有一个由 4 个碱基 T 组成的环和一段位于 3′端的 8 个碱基 T，因此与大肠杆菌的转录终止信号非常相似。

这三种噬菌体的基因组都含有一个以上转录终止子。f1 噬菌体的终止子 I 和终止子Ⅱ的差别主要在于后者有一个非正规的 G－T 碱基对，主杆中有一个未配对的碱基 A 以及 3′端 T 碱基串的差别。M13 的终止子 I 与终止子Ⅱ则毫无同源性，后者 3′端也不存在 T 碱基串。

表 6－5　噬菌体基因的转录终止信号

噬菌体 T7						
TE	CGTTTATAAGGA－	GACACTTTATGT	TTAA	GAAGGTTGG	TAAATT－C	CTTGCGGCTTTG
TΦ	TGCTGA－AAGGAG	GAACTATATGCG	CTCA	TACGATATG	AACGTTGA	GACTGCCGCTGA
噬菌体 f1						
I	TAAACCGATACA	ATT*AAAGGCTCC*	TTTT	*GGAGCCTTT*	TTTTTTGG	AGATTTTCAAC－
II	CCCTTTGACGTT	*GGAGTCCACG * TT*	CTTT	*AAT * AGTGGA*	CTCTTGTT	CCAAACTGGAAC
噬菌体 fd						
I	TAAACCGATACA	ATT*AAAGGCTCC*	TTTT	*GGAGCCTTT*	TTTTTTGG	AGATTTTCAAC－
噬菌体 M13						
I	TAAACCGATACA	ATT*AAAGGCTCC*	TTTT	*GGAGCCTTT*	TTTTT	
II		*AACCTCCCG*	CAAG	*TCGGGAGGT*	TCGCT	
噬菌体 IKe						
I	TTTTCAGCGTTA	TTT*AAAGGGGCGC*	TATT	*GCGCCCCTTT*	TTTTTACT	TTAATTCAGCTA
噬菌体 ΦX174						
T4	GTATGT	TTTCAT*GCCTCC*	*AAA*T	*CTTGGAGGC*	*TT*TTTTAT	GGTTCGTTCT－－
T2	CAACAATTTTAA	TTGC*AGGGGC*TT	CG*GC*	*CCCT*TACTT	GAGGAT	
T1	ACTATA	GACCACCGCCCC	GAAC	GGGACGAAA	AATGGTTT	TTAGAGAACG
噬菌体 P22						
T ant	GATAACCAAC	GCAAC*GACCCAG*	CTTC	G*GCTGGGTT*	TTTTTATG	
T nutR	CCA	ATCTGA*ACCGCC*	GACA	AC*CCGGT*AA	ACC	
R1		TCAAAG*CGCA*	*T*－CA	ACGA*ATGCG*	CACACTA	ACTAT
噬菌体 λ						
T nutR		*GCCCTG*	AAAA	*AGGGC*ATCA	AATTAAAC	CACAC
R1		CTATGG*TGTATG*	*CA*TT	TATTTGCAT	ACATTCAA	TCAATT
噬菌体 Pf3						
CP		TCGTT*ATAAGG*	*GGGC*	TTC*GGCTCC*	*CTTAT*TCG	TTTA

表 6－5 中最后一种病毒 pf3 为假单孢菌（*Pseudomonas aeruginosa*）的噬菌体，其转录终止信号与上述噬菌体基因的终止子的碱基序列毫无相似之处，但它具有一个相似的双重对称结构，其中有 10 对配对的碱基，包括一对非正规的 TG 碱基对。此外，这种结构的下游有 5 个碱基 T，但这段序列中间插有 CG 两个碱基。

3. 其他噬菌体的终止子　一般情况下，同一类型噬菌体的终止子是相似的。例如 φX174 的终止子 4（T4）与上述 4 种丝状噬菌体的终止子都非常相似，终止子 T4 的对称结构为 GCCTCC 和 GGAGGC，而 f1 噬菌体终止子 I 的对称结构为 GGCTCC 和 GGAGCC，并且二者之后都有一段 T 碱基串。但是并非所有终止子都是如此，如终止子 T2 就与 f1 噬菌体的终止子毫无相似之处。

在 P22 和 λ 噬菌体之间，终止子也同样存在结构差异。这两种噬菌体的 nutR 基因终止子的主杆都很短，只有 5 对碱基，所以存在一定相似性。但这两种噬菌体的 R1 终止子无论在长度或结构方面都存在很大差异。P22 的另一基因 ant 的终止子的主杆则与 nutR 基因的终止子具有序列同源性。

（二）真核生物病毒基因的终止子

在脊椎动物病毒中，转录的终止信号目前还不清楚，其主要原因是大多数 DNA 病毒的基因组均为环状分子，转录通常是双向进行，而且早、晚期基因的表达共用转录启始和终止区。根据对 SV_{40} 的研究，多聚 A 化信号序列 AAUAAA 除了具有使 mRNA 在转录后加尾的作用外，还具有终止转录的功能。

在植物病毒中，根据对烟草花叶病毒（TMV）的西红柿病毒株以及黄瓜花叶病毒的研究，其 RNA2 不具有在典型的多聚 A 化信号序列，但在其他几条基因组 RNA 中则存在这种加尾信号。在烟草条纹病毒（TSV）中，其基因组 RNA3 不含这种信号序列，但其 5′ 端的前序列中有 3 段序列重复。这些重复序列的确切功能还不清楚，它们可能具有终止子的功能。此外，某些植物病毒基因组的转录终止信号位于 tRNA 顺反子中，可能这种 tRNA 序列本身就具有终止转录的功能。

三、RNA 病毒蛋白质合成的策略

所有由病毒基因组指导合成的蛋白质都是在宿主细胞质核糖体上合成的。虽然各种病毒的翻译机制基本相似，但不同病毒在蛋白质合成的策略上有明显不同。如前述脊髓灰质炎病毒将其单正链 RNA 基因组翻译成一条巨大的多蛋白质分子，然后再将多蛋白分子加工成几条较短的多肽。又如，烟草叶脉斑点病毒（tobacco vein mottle virus，TVMV）也是首先将其单正链 RNA 基因组翻

译成一条大分子蛋白质，然后再加工成几条较小的蛋白质（图 6－28B）。

RNA 病毒基因组表达的第二种策略是以病毒基因组为模板，合成多条 mRNA 分子。如前述流感病毒基因组由 8 条不同的单负链 RNA 分子组成，每一条基因组 RNA 均为一种 mRNA 的模板，然后各条 mRNA 再翻译成不同的病毒多肽。又如雀麦花叶病毒（brome mosaic virus）的三条单正链基因组 RNA 也采用相似的策略进行表达（图 6－28），惟一的差别是在基因表达过程中无须进行转录。但是，雀麦花叶病毒的外壳蛋白质基因位于二顺反子 RNA 上，在表达过程中可能需要通过复制与另一顺反子分离。

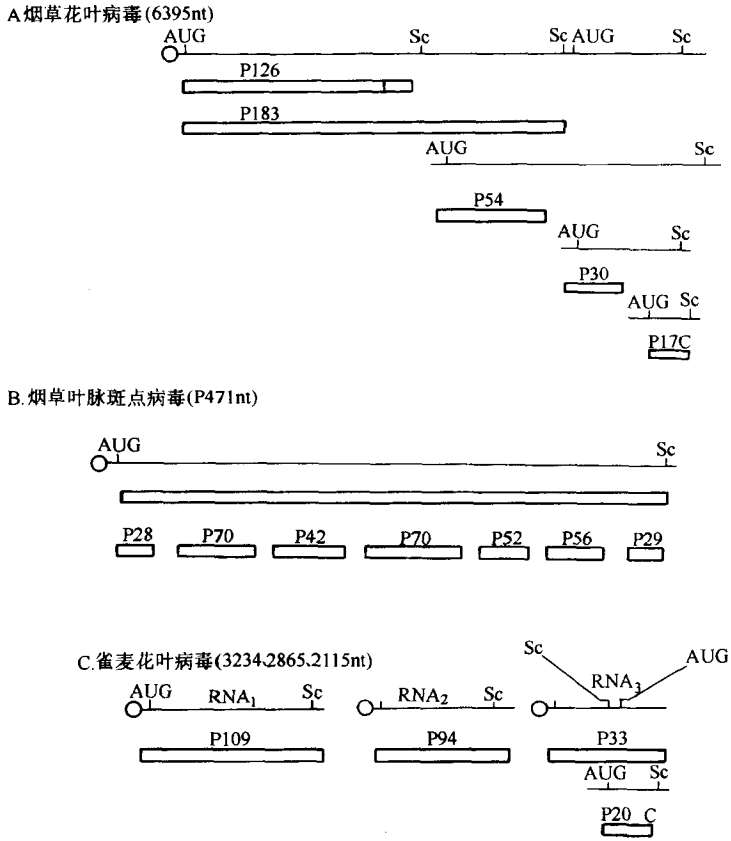


图 6－28　RNA 病毒基因组表达的三种主要策略

第三种表达方式是病毒基因组 RNA 首先转录成若干个单顺反子 mRNA 片段，然后再各自进行翻译。如烟草花叶病毒可能首先通过复制过程进行基因组分割，产生亚基因组 RNA，然后再翻译成不同蛋白质（6－28A）。烟草花叶

病毒也可同时采用另一种不同方式进行表达，如其5′顺反子的末端有一个终止密码，当阅读通过这一密码时，可以产生两种氨基末端氨基酸序列相同的蛋白质。

动物的某些单正链RNA病毒也可采用相似策略进行基因表达。例如披膜病毒（toga－ virus）的基因组为一条二顺反子正链RNA，其5′端的一个顺反子编码病毒复制酶，3′端的顺反子编码病毒粒子蛋白的前体（图6－29）。由于宿主细胞的核糖体不能识别基因组内部的翻译启始位点，所以基因组本身只能用作复制酶蛋白mRNA，即只有5′顺反子才能翻译。壳体蛋白只能在病毒感染的后期合成，它是由基因组3′端的亚基因组mRNA翻译而成。这种mRNA是通过在负链的内部位点上启始复制而形成的。由于mRNA不具有复制酶顺反子，所以壳体蛋白的翻译起点密码就能暴露给核糖体，用于翻译启始过程。

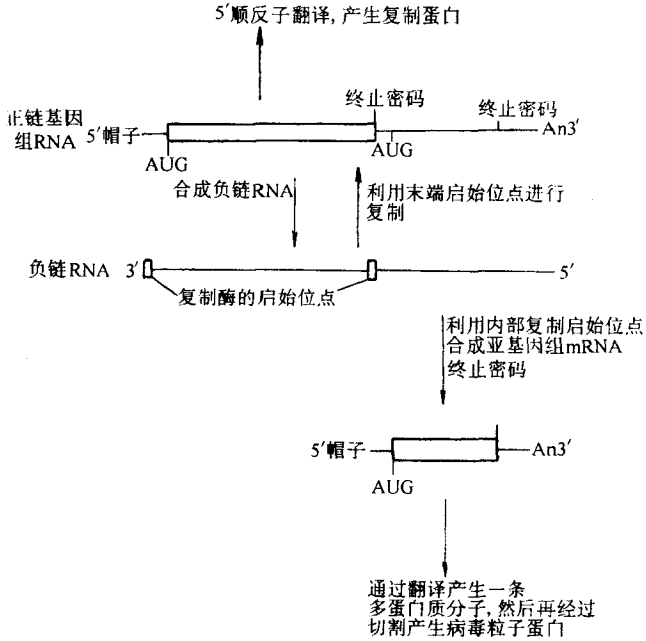


图6－29　披膜病毒基因组的表达

在感染后的早期，正链RNA基因组通过翻译，产生两种非结构蛋白，其中包括复制酶。在感染的后期，利用负链上的内部复制启始信号，合成一条亚基因组mRNA。亚基因组翻译成一条前体多肽，然后前体多肽通过切割，形成病毒粒子蛋白，包括壳体蛋白和三种与质膜结合的糖蛋白

在一种单正链RNA病毒中，同一条RNA分子既可用作蛋白质合成的信使，同时也可用作基因组物质。但是同一条RNA链不能同时进行翻译和复

制，RNA 只有脱去核糖体以后才能进行复制。可能这些病毒都利用蛋白质来抑制其翻译。例如在噬菌体 ms2 和 Qβ 中，当壳体蛋白与 RNA 某个区域结合后，就会阻止复制酶基因的翻译，而复制酶复合体与 RNA 结合后，就可抑制壳体蛋白的翻译。

第四节　病毒与肿瘤

肿瘤是一大类严重威胁人类健康并危及人类生命的疾病。全世界每年约有 870 万人患肿瘤，死亡人数约 690 万。成为疾病死亡原因的第二位（仅次于心血管疾病）。

一、肿瘤及癌的一般生物学概念

肿瘤是一种异质性的细胞异常增生性疾病。肿癌细胞的异质性是指其在结构、功能、代谢和行为上都已不同于正常细胞。上皮细胞的恶性肿瘤称为癌，约占全部恶性肿瘤的 90% 以上。癌细胞是一种失去接触抑制的低分化性和获得永生性的细胞。正常细胞在化学、物理、生物等致瘤因素作用下，其 DNA 的结构发生改变和基因表达异常就会形成变异细胞。变异细胞或者逆转为正常细胞，或者在某些因素的进一步作用下形成癌细胞。癌细胞在逃逸了宿主免疫监视和抑制了宿主的生长限制之后，就会增生为癌细胞克隆并演进为具有异质性的多种亚克隆，进而形成癌肿。失控生长的癌肿从原发位向周围组织侵袭或远处扩散，这种侵袭或扩散若无法控制将很快侵犯要害器官，引起器官衰竭，最终导致机体死亡。

肿瘤有良性肿瘤和恶性肿瘤之分，良性肿瘤的特征是：(1) 细胞被正常组织所局限，常有一层纤维结缔组织包围；(2) 生长缓慢，呈膨胀性，对周围组织不具侵袭破坏性；(3) 不发生远处转移。恶性肿瘤的特征是：①细胞呈侵润性生长，周围无结缔组织包膜；②生长速度快，侵袭和破坏周围正常组织；③向远处转移和播散，形成继发性肿瘤。

目前普遍认为，癌细胞与正常细胞在本质上的变异主要表现在 2 个方面：(1) 癌细胞在细胞增殖和细胞分化 2 个生物学过程之间的调节控制上异常，癌细胞不断生长、分裂，却不发生功能分化；(2) 癌细胞是变异细胞，其变异性从分子上可表现为 DNA、RNA、蛋白质等构成细胞的大分子的变异；从细胞上可表现为染色体、核型、微管等方面的变异；从生长上，可表现为生长速度、分化速度、营养要求、接触抑制和浸润转移等行为方面的变异。癌细胞的变异性构

成了癌细胞的异质性，其本质在于癌细胞遗传的不稳定性和基因突变。

总之，癌的发生是一个多因素、多基因、多步骤的复杂的生物学过程。癌的自然发展史，可以分为启动、促进和演进3个不同又连续的过程。癌的病因涉及环境、遗传、免疫、营养方面的问题。环境中致癌因素，主要包括化学因素、物理因素、生物因素、遗传因素等能够引起遗传物质DNA的结构和功能改变。但癌变涉及细胞的基本生物学规律的改变。因此，从细胞的遗传、变异、发育、分化、代偿、老化、凋亡等基本生物学规律入手，研究癌的发生与发展规律已成为新的研究热点。癌基因、抑癌基因、转移基因和转移抑制基因等癌变过程中的关键性基因以及影响肿瘤生长的各种生长因子、生长抑制因子的研究不断深入。细胞的分化、去分化、肿瘤细胞的浸润性生长，肿瘤血管形成、肿瘤逆转，癌变过程中启动、促进和演化的机制也都处在深入研究之中，这无疑将更加深化了人类对癌起因的本质问题的认识。

二、动物及人体的致癌因子

肿瘤病因学研究的对象是引起肿瘤发生的始动因素，这些因素包括化学因素、物理因素、生物因素、遗传因素等。

（一）化学因素

据世界卫生组织（WHO）发表的资料表明，人类恶性肿瘤80%～90%与化学因素有关。化学物质在人类恶性肿瘤发生中的特殊地位，引起了人们对化学致癌物进行实验研究的重视，确切地认识某些化学致瘤物与恶性肿瘤的关系，就可以在人们大规模地接触之前就能确立其致癌性，为早期预防提供依据。

1. *化学致癌物种类*　烷化剂类　有致癌作用的烷化剂有：氮芥和硫芥类、乙撑亚胺类、磺酸酯类以及常被用来作为化疗剂、灭菌剂、杀虫剂等的一些化学物质。临床使用这类药物治疗一种癌症，患者往往又因此导致第2种癌症。致癌性杀虫剂在农作物中的残留及对环境的污染所导致的对人类的潜在性危害也引起了重视。

多环芳烃类化合物　是指由多个苯环缩合而成的化合物及其衍生物，其代表性物质有3，4－苯并芘、二甲基苯蒽、二苯蒽，3－甲基胆蒽等，燃烧的卷烟、不完全燃烧的脂肪、煤炭、石油以及用烟熏制的肉类食品等均可产生多环芳烃类化合物。

芳香胺类化合物　主要有芳香胺类染料和芳香酰胺类化合物，如：γ－萘胺、联苯胺、4－氨基联苯、4－硝基联苯、联甲苯胺、金胺、邻联茴香胺等。

其中以γ—萘胺的致癌性最强。

氨基偶氮类染料　这类化合物含有偶氮基团－N＝N－，因其具有颜色，常被用作纺织品、食品及饮料等物品的染料或添加剂。给实验动物长期、大剂量喂养时可诱发肿瘤。

亚硝胺类化合物　是一类作用很强的环境化学致癌物，如二乙基亚硝胺、甲基苄胺、哌啶、肌氨酸、乙酯亚硝胺、甲基亚硝基脲烷等。这类致癌物主要存在于食物及吸烟的烟雾中。一般说来，腌制食品、剩饭菜及腐败蔬菜中含量多。

植物毒素中有苏铁素、蕨毒素、黄梓素等可诱发肝癌、膀胱癌及食管癌。

2. 化学致癌物的生物学效应　致癌物经体内酶系统代谢活化以后，形成具有致癌性的终致癌物，终致癌物与细胞 DNA 结合，形成加成物，并使其烷基化。此外终致癌物也可与 RNA、蛋白质产生烷基化反应，DNA 上有多个位点可发生烷基化，它们是 N－7、O－6、N－3、胞嘧啶和磷酸三酯。DNA 加成物是机体接触化学致癌物的标志，致癌物剂量越大，加成物水平越高，则肿瘤发病率越高，而且还可以缩短潜伏期。

3. 促癌物及其作用　在对兔子进行煤焦油致癌实验时发现，在兔子皮肤上反复使用煤焦油可引起肿瘤，但如果停止使用后肿瘤便消退。若再用无致癌作用的松油刺激皮肤，可使肿瘤重新出现。这一现象提示：煤焦油和松油引起 2 种不同的作用，即启始作用和促动作用。启始作用是将正常细胞不可逆地转化为肿瘤细胞，又经细胞分裂增殖而固定下来，并可传给子代细胞，从而启动了细胞的癌变。这一阶段的特点是作用快，过程短、不可逆，细胞 DNA 产生永久性变化。促动作用是化学物质通过作用于细胞膜，使其产生系列化改变而促进细胞的分裂，其特点是作用慢，所需时间长，有可逆性。

具有促癌作用的物质主要有多种工业来源的污染物、植物来源以及内源性促癌物等。工业来源的促癌物以蒽酚最为活泼，烷基苯磺酸盐（用于去垢剂）具有明显促癌作用。植物来源的促癌物主要有巴豆油（crotonoil），其有效成分为一组佛浓醇酯类。内源性促癌物主要是激素，目前已发现性激素、肾上腺素、肾上腺皮质激素及生长激素等与外源性致癌因子产生协同作用，促进肿瘤的发生。

促癌物与致癌物虽然在致癌过程中都发挥重要的作用，但它们的生物学作用特点又各不相同，二者的区别如表 6－6。促癌物主要通过促进细胞增殖，促进靶细胞 DNA、RNA 及蛋白质的合成，增强各类致癌病毒转化细胞的能力，促进癌基因的表达。促癌物对已启动癌变的细胞可产生抑制分化的作用，使其逐步演化为癌细胞。

表 6-6　促癌物与致癌物的区别

促　癌　物	致　癌　物
1. 单独使用无致癌性，须由致癌物作用后才发挥作用	1. 本身具有致癌性，常单一剂量便可启始癌变
2. 不具致突变性，但可促使已突变的细胞表达	2. 多为诱变剂，可产生明显的诱变效应
3. 作用是可逆的，不积累，须重复作用	3. 作用不可逆，可积累，无剂量阈值。
4. 无需代谢活化或与大分子结合	4. 多数须代谢活化，并与大分子共价结合。
5. 引起细胞生物学改变是渐进性的，呈现恶化前可见到各个稳定的过渡阶段	5. 引起细胞生物学改变的作用快，时间短。

（二）物理因素

所谓物理致癌因素主要是指辐射即电离辐射和非电离辐射，包括γ射线、X射线、宇宙射线和放射性核元素以及紫外线。随着科学技术的发展，电离辐射和紫外线被广泛用于工业、农业、能源、国防与医疗等领域，为人类物质文明做出了重要贡献，但随之而来的对环境的污染及对人类健康的危害，尤其是其远期效应，也越来越受到人们的广泛关注。实际上，辐射致癌已成为人类肿瘤的重要病因之一。

1. 紫外线辐射与癌　阳光中，波长在290nm以下的短波紫外线被大气臭氧层吸收，290nm以上的长波紫外线可达到地球，紫外线可透过人的机体皮肤表层到达真皮表层，若长期曝晒，可使皮肤老化，添发各种不同的皮肤肿瘤，主要有基底细胞瘤和鳞状细胞瘤。已证实，由于紫外线可被正常的皮肤色素所吸收，因此，正常肤色黑的人患皮肤癌的概率比肤色白的人要低。

紫外线穿透力不强，所致癌多局限于皮肤。DNA接受紫外光后，2个相邻的嘧啶碱基产生反应，形成嘧啶二聚体，导致DNA损伤，并抑DNA复制。这种损伤若不能及时修复时或修复功能不全时，则可导致皮肤细胞发生癌变。

2. 电离辐射与癌　电离辐射主要包括X射线、γ射线和宇宙射线等，其特点是波长短，能量强，足以穿透机体深部组织。沿着这条能量穿过的轨迹，分子可出现电离。目前已肯定接触电离辐射能增加肿瘤发病率，如接受X－射线检查等。放射也用于10多种恶性肿瘤的治疗，但由此也会引发继发癌问题。日本遭原子弹爆炸后2～3年，在曾暴露于核辐射的人群中，首先出现恶性肿瘤发病率升高的是白血病，其潜伏期较其他肿瘤短，10年左右达到最高峰，白血病类型出现频率高低依次为急性粒细胞白血病、慢性粒细胞白血病、急性淋巴细胞白血病。此外，在这类人群中乳腺癌、甲状腺癌、肺癌以及其他部位的肿瘤的发病率亦上升。

辐射主要作用于癌变过程中的早期阶段，正常细胞 DNA 中含有无致癌活性的原癌基因，在辐射作用下，原癌基因结构发生变化而导致异常表达，使其转化为具有致癌活性的癌基因。正常细胞由此获得新的遗传信息，并经过一系列有关酶、大分子物质的变化及相互作用，便启动细胞癌变。此外，辐射造成染色体畸变，使 DNA 单链或双链断裂，DNA 修复系统功能出现错误，也会启动细胞转化。

（三）病毒因素

病毒主要靠将其基因整合到宿主细胞 DNA 中去改变细胞 DNA 而发挥致癌作用。病毒因素致癌过程以及参与病毒因素致病过程中的肿瘤相关基因及作用方式如图 6－30 所示。

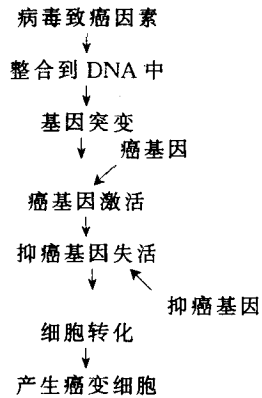


图 6－30　病毒因素致癌过程中肿瘤相关基因的参与

按照肿瘤病毒所含核酸的不同，分为 DNA 肿瘤病毒和 RNA 肿瘤病毒。RNA 肿瘤病毒都是逆转录病毒，在自然界广泛分布，对爬虫类、禽类、哺乳类和灵长类有致瘤作用，可诱发这些动物的肉瘤、白血病、淋巴瘤及乳腺癌等。RNA 肿瘤病毒共有的特征是，病毒颗粒中都含有逆转录酶，所以在分类上属于逆转录病毒。形态及结构都基本相似，多呈圆形或类圆形，成熟的病毒颗粒有一个电子密度较深的核心，外包以核壳（核衣壳），最外层还有一层囊膜包绕。

按在电子显微镜下的形态，RNA 肿瘤病毒可分为 4 种不同的类型。

A 型病毒　具有两层电子密度较深而分界不甚清晰的包膜，中央部分为电子密度较淡的透亮区，直径约 70～90nm，分布在胞质内或扩大的囊泡内，这类病毒一般被认为是不成熟的病毒颗粒。

B 型病毒　具有偏位的类核，包膜表面有明显的刺突（Spike），其典型代表病毒为小鼠乳腺肿瘤病毒。

C 型病毒　成熟的病毒颗粒中心有密度较深的类核，包膜与类核间为电子密度较淡的晕。该类病毒是 RNA 肿瘤病毒中最常见的一类，与小鼠、大鼠、猪、牛、猴等的白血病、肉瘤和淋巴瘤发生有关。

D 型病毒　与 B 型类似，有电子密度较深的偏位的小棒状类核，病毒的外形类似希腊字母 θ，属于这类病毒的有猴乳腺肿瘤病毒和人类免疫缺陷病毒。

DNA 肿瘤病毒是指一类在自然界能感染动物或人、并在体内能诱发肿瘤，在体外可诱导正常细胞转化、所含核酸为 DNA 的病毒。它分属于 DNA 病毒的多瘤病毒科、乳头状病毒科、腺病毒科、疱疹病毒科和嗜肝病毒科（表 6－7）。DNA 肿瘤病毒在大小和形态方面差别很大，但其超微结构相似，皆为 20 面立体对称，除疱疹病毒外均无包膜（envelope）。病毒核心（core）只是 DNA 和蛋白质组成的复合体。外有衣壳（capsid）包绕，衣壳又由数目不等的壳粒（capsomere）组成。

表 6－7　DNA 肿瘤病毒

科	主要代表病毒	科	主要代表病毒
多瘤病毒科	SV_{40}、多瘤病毒、	疱疹病毒科	人单纯疱疹病毒Ⅰ型
	BK 病毒、JC 病毒		人单纯疱疹病毒Ⅱ型
乳头瘤病毒科	人乳头瘤病毒	马疱疹病毒Ⅰ型	
	shope 乳头病毒		人巨细胞病毒
	牛乳头瘤病毒		EB 病毒
	其它动物乳头瘤病毒		马立克病毒
			猴疱疹病毒
			Lucke 病毒
			棉底兔疱疹病毒
腺病毒科	人腺病毒，至少 31 个亚型	嗜肝病毒科	乙型肝炎病毒
	动物腺病毒		土拨鼠肝炎病毒
	猴腺病毒，24 个亚型		地松鼠肝炎病毒
	牛腺病毒，9 个亚型		鸭肝炎病毒
	猪腺病毒，14 个亚型		
	其他动物（羊、狗、小鼠、马、禽类）、腺病毒		

三、癌基因、抑癌基因及肿瘤转移调节基因

（一）癌基因

1969 年，Huebner 和 Todaro 在劳氏肉瘤病毒基因组中发现了一个与癌变相关的基因，即病毒癌基因（V－oncogene）。1976 年，美国加州大学 Bishop

和 Varmus 证实了病毒癌基因可以引起细胞转化。目前已分离、鉴定了数十种病毒的致癌基因，而且已证明脊椎动物细胞都存在这些致癌基因，人体细胞内也含有这些基因，这种基因称为细胞癌基因（cellular oncogene）。存在于正常细胞中未被激活的癌基因称为原癌基因（proto－oncogene），当被某些物理、化学或生物因素刺激后即被激活的原癌基因即为癌基因。细胞中有阻遏物存在，在正常情况下癌基因不表达，由于射线、致癌物、肿癌病毒和衰老过程都可能使阻遏物活性减退，使其产生转化蛋白，使正常细胞转化为癌细胞。

1．病毒癌基因（V－onc） 现已发现了 30 种以上的 V－onc 基因，并对它们的结构和编码产物的性质、功能等方面进行了深入的研究。病毒癌基因分为 RNA 病毒中的 V－onc 和 DNA 病毒中的 V－onc，但一般是指急性转化型逆转录病毒中的癌基因，包括 src 基因家族、sis 基因家族、ras 基因家族、myc 基因家族、myb 基因家族以及其他基因。V－sis 基因编码的 P28V－SIS 与人血小板生长因子（PDGF）结构非常类似，PDGF 能刺激间叶组织的细胞分裂，它与细胞相应的受体结合后，通过受体自身的酪氨酸蛋白激酶使细胞内多种蛋白质磷酸化，以发挥其促进 DNA 合成、加速多种代谢过程、调节细胞生长等作用。src 基因编码与细胞恶性转化关系密切的酪氨酸蛋白激酶；Ras 基因编码的蛋白质为 P21ras，与 GTP 或 GDP 有高度亲和力，异常的 P21 易被 GTP 活化成为具有致瘤能力的转化蛋白。myc 基因、myb 基因编码的蛋白是能与 DNA 结合的核蛋白，当细胞受到促分裂刺激后，这类基因的表达急剧增加。

2．细胞癌基因及其产物

（1）细胞癌基因（c－onc）具有以下特点

① c－onc 与 v－onc 是同源物。目前认为，RNA 致瘤病毒中的 v－onc 是在长期生物演化过程中，病毒从宿主细胞中摄取的 c－onc 经过加工后成为致瘤病毒的 v－onc 基因。动物致瘤病毒的研究导致人的 c－onc 的发现，这是病毒肿瘤学的重要贡献。

②c－onc 不仅存在于灵长类和人类，而且也存在于果蝇和酵母，从种系进化上看，前后相关十亿年之久。c－onc 在这么漫长的进化过程中保持下来，必有其重要的生理功能，它在细胞分化过程中可能占有独特的地位。

③一般认为，在正常情况下，c－onc 处于静止状态，并不对机体造成危害，所以遗传学中常将细胞原癌基因与 c－onc 混用。

④目前发现的 c－onc 都是单拷贝基因，并且都是结构基因，因此都有编码产物。c－onc 通过其产物引起细胞增殖。

⑤ 没有严格的种属特异性和组织特异性，活化的c－onc可在不同的动物

表 6－8　人 c－onc 定位及产物

c－onc 基因	产　　物	产物定位	生物活化
sis	DPGF 的 β 链	膜外	生长因子
src	pp60c－src	质膜、核膜	酪氨酸磷酸化 InsP3＋DG
fps	p98c－fps	质膜、胞质	酪氨酸磷酸化
fes	p92c－fes	质膜	同上
abl	p150c－abl	质膜	同上
yes	?	?	同上
fms	?	内质网、质膜	同上
mos	?	胞质	丝/酪磷酸化
ras	?	质膜	InsP3＋DG
raf－1	?	胞质	丝氨酸磷酸化
raf－2	?	胞质	同上
erb－B	切短的 EGF	内质网、质膜	酪氨酸磷酸化
Ha－ras－1	p21c－Ha－ras－1	质膜	结合 GTP 或 GDP
Ha－ras－2	p21c－Ha－ras－2	质膜	同上
Ki－ras－1	p21c－Ki－ras－2	质膜	同上
Ki－ras－2	p21c－Ki－ras－2	质膜	同上
N－ras	p21c－N－ras	质膜	同上
mys	p21c－mys	核内与 DNA 结合	丝/酪磷酸化
L－myc			
N－myc			
myb	?	核内	
fos	p55c－fos	核内	
Blym	转铁蛋白	核内	
ski			

与不同的组织内表达。

（2）癌基因产物的分类　现已发现了 30 多种 c－onc。其产物按其功能大致可分为：生长因子类、膜受体类、传导分子类、核内蛋白类以及其他类等。

已经发现的属于生长因子和生长因子受体类的癌基因有编码人血小板生长

因子（PDGF）β链的 sis 癌基因，编码神经生长因子（NGF）受体的 trk 癌基因，编码上皮生长因子（FGF）受体的 erb－B 癌基因，编码肝细胞生长因子（HGF）受体的 met 癌基因，编码集落刺激因子（CGF－1）受体的 fms 癌基因及编码胰岛素样生长因子（IGF）的癌基因等类型（表 6－8）。它们是编码生长因子或者编码生长因子的受体。属于蛋白激酶的癌基因可再分为 3 个亚类：①具有受体同时具有酪氨酸激酶功能的癌基因。例如编码 EGFR 受体的癌基因、编码纤维细胞生长因子受体的癌基因以及编码膜外区有重复序列和编码膜外 2 个类似免疫球蛋白区的癌基因等都属此类。②不具有受体功能而仅具有酪氨酸激酶功能的癌基因。例如 C－src、Lak、C－fgr、hcl、blk、C－yes、lyn、fyn、src 癌基因家族成员以及 C－abl、C－fes、sak、atk 等癌基因也属此列。③具有丝氨酸和苏氨酸激酶活性的癌基因。这类癌基因主要有 C－raf－1、A－raf、B－raf、mos、pim－1、bcr 等。属于具有 G 蛋白功能的癌基因是一大类，主要由 ras 家族组成。包括 H－ras、K－ras（A、B）、N－ras、raplS（krev－1）、rap1B、rap2、R－ras、ralA、ralB、TC21、rhoA、thoB、rhoC、tacl（TC25）、rac2、rab1A、rab1B、rab2、rab3A 等基因。属于核内癌基因类型的有 C－ros/C－jun 家族、C－myc 家族、C－erbA 家族、C－rel、C－myb 家族、ets 家族、ski 等癌基因。这类癌基因编码的蛋白质多为转录相关的因子，如转录因子、转录调节因子、转录启动因子、转录促进因子等。

（二）癌基因产物的作用机制

（1）生长因子及受体类癌基因产物　生长因子与受体结合使受体自身被激活是受体开始发挥作用的第一步。癌基因编码的生长因子受体的激活也同样如此。如 PDGF 受体激活后至少可以与 4 种不同的蛋白结合，形成一个大的复合体促使受体进一步发挥作用。

（2）蛋白质激酶类癌基因产物　许多蛋白质在磷酸激酶的作用下加上磷酸（磷酸化）时具有了活性，而去磷酸后就失去了活性。磷酸化与去磷酸化就像生命开关一样在调节着蛋白质的功能。对癌基因产物的研究发现，大约 50% 的蛋白质激酶是原癌基因或癌基因的产物。蛋白质激酶的活化是癌基因产物表现其功能的关键一步。

（3）G 蛋白类癌基因产物　主要是 ras 癌基因以及 ras 相关基因编码的一类蛋白质。ras 癌基因蛋白的特征是它们与典型的 G 蛋白一样能与 GTP 和 GDP 结合。它们与 GTP 结合后才有活性。还有 2 类调节因子参与了 ras 蛋白的 GAP 蛋白的活化；另一类是 GTP/GDP 交换因子，前者可与正常的 ras 原癌基因蛋白结合，提高 GTP 酶活性至少 100 倍。GAP 蛋白也可以与突变的 ras 蛋白结合，但这种结合不能激活 GTP 酶，因此认为 GAP 蛋白是一种负性调节

因子。GAP 蛋白高效表达可以阻止 C－H－ras 基因表达引起的细胞转化。GTP/GDP 交换因子是一种正调节因子，其作用机制是促进 ras 蛋白上的 GDP 释放，从而使 ras 蛋白与 GTP 结合处于活性状态，激活 ras 蛋白释放 GDP 和 GTP，使 ras 蛋白活化后表现出 G 蛋白活性。ras 蛋白活化后即可与效应因子结合，促进细胞增殖。

（4）核内癌基因产物　核内癌基因大都能表达相关蛋白，如转录因子、转录调节因子、DNA 结合蛋白等。生长因子、细胞因子、促癌剂、致癌剂都能改变转录因子（AP－1）的转录活性；转录调节蛋白具有识别和与特异 DNA 序列结合的功能；p50－p65 二聚体是 rel 家族中的一个成员，可识别 DNA 的 5′－GGGACTTTCC－3′序列，p50/p65 二聚体处于静止状态时，位于细胞质内，这时一个抑制蛋白 LKB 与之结合使其处于静止状态。当细胞接受细胞因子作用，激活 PKC 后，PKC 与 LKB 结合，释放出 p50－p65 蛋白，并随即转入核内与 DNA 结合，促进基因的转录。

（三）细胞癌基因的激活机制

c－onc 属于正常的细胞结构基因之一，与细胞的生长、分化、代谢等有关。虽然 c－onc 在正常情况下也能表达，但相对说来仍处于抑制状态，在一定条件下被激活之后才能导致肿瘤的发生。因此，c－onc 的激活是肿瘤发生过程中的一个关键步骤。不同的致瘤因素通过不同的机制使 c－onc 活化，同一种 c－onc 的活化在致瘤过程中常涉及不同的激活机制，但同一致瘤因素可以通过不同机制、不同致瘤因素也可通过同一机制激活 c－onc。常见的激活机制可以概括为以下几个方面。

1．*原癌基因获得活化因子*

①原病毒的插入：禽类白细胞增多症病毒（ALV）感染细胞后，可以激活细胞的 c－myc 原癌基因，诱发鸡 B 淋巴细胞瘤。ALV 本身并无转化基因，但其病毒两端的长末端重复序列（LTR）含有启动子，插入到细胞 c－myc 基因附近适当部位后，可使 c－myc 活化，使 c－myc 专一的 RNA 拷贝增加近 100 倍。进一步研究发现，病毒 LTR 对 c－onc 基因的激活有多种形式。在上述 B 淋巴细胞瘤中，LTR 可以插入到 c－myc 上方，并与细胞基因同方向，直接驱动 c－myc 转录，也可以逆向插入到 c－myc 上方或顺向插入 c－myc 下方，增强 c－myc 基因本身的转录。病毒序列插入远离 c－onc 的位点也能激活 c－onc。

②染色体易位与基因重排：c－onc 的编码序列（即外显子）在 DNA 上不是连续的，内含子起着促进和调节基因的表达作用。例如人 c－fos 和 c－Ha－ras－1 都是有 4 个外显子和 3 个内含子组成，人 c－myc 由 3 个外显子和 2 个

内含子组成。不具有这种“外显子－内含子”结构的c－onc为数极少。因此，如果染色体出现易位、交换等变异时，可使c－onc与正常的抑制子分开而被置于活化DNA序列的控制之下，因而被激活。逆转录病毒具有一种与细胞转座子类似的结构，而染色体上则有一些与c－onc密切相关的脆弱位点，它们易受致癌因素的作用而断裂，使染色体发生易位，导致c－onc重排。

③染色体非组蛋白改变：非组蛋白是染色质的构成分子之一，与组蛋白不同，它既有种族特异性，又有组织特异性，对一般的细胞基因及c－onc的表达起重要的调控作用。非组蛋白的增加与启动细胞恶变所必需的染色质变化有关。

2．原癌基因抑制的清除

①c－onc甲基化水平降低：DNA分子的甲基化能增加DNA分子双螺旋的稳定性，抑制启动子和转录过程，在基因活化过程中起重要作用。某些化学致癌物的致癌性也与它们抑制或灭活甲基转移酶、阻止S－腺苷蛋氨酸与DNA分子之间甲基转移过程的能力有关。一些致癌剂可使细胞内源c－onc去甲基化而被活化，最终导致癌变。病毒引起转化的细胞也表现出基因甲基化水平降低。

②c－onc“逃脱”抑制因子的作用：研究发现，c－myc的第一个外显子不编码，它可能有抑制c－myc重排的作用。当染色体易位或病毒感染使c－myc重排或被修饰，这个外显子被剪切后，c－myc摆脱正常情况下所受的抑制，才能受增强子的作用被激活。

3．原癌基因突变　在同一个c－onc中使原癌基因活化的点突变可有多种形式，同一位点上可以出现不同的突变，即不同的核苷酸可以取代同一位点上正常的核甘酸，同一密码子内不同的位点上或不同的密码子内也可以出现这种突变。

在人c－onc中，一个核苷酸的突变就能使c－onc被激活、异常表达和使正常细胞转化。如与膀胱癌有关的基因是C－Ha－ras－1，该基因在正常细胞和肿瘤细胞中的差异在于它第一个外显子第35位（12号密码子）上的脱氧鸟嘌呤核苷酸被胸腺嘧啶核苷酸取代（G→T），原癌基因因此而被激活，结果使正常的p21蛋白质第12位上的甘氨酸被缬氨酸取代，进而使细胞恶变。

4．原癌基因扩增　某些原癌基因复制时可以由一个拷贝扩增为多个拷贝。原癌基因拷贝数的增加会导致基因产物增加，从而引起细胞正常功能紊乱。在一个癌细胞中，一般不止一种原癌基因被激活，往往多个原癌基因在肿癌发生的多个阶段上，相继或同时被激活。目前认为，在癌变过程中至少有2类原癌基因被激活才能完成癌变过程。一类是使细胞产生永生性的癌基因，如myc

癌基因等。另一类是使细胞迅速增殖、细胞表面形态和功能改变的癌基因，如ras癌基因，在致癌过程中myc和ras癌基因互补才能使细胞恶变。

（四）抑癌基因

抑癌基因（tumor suppressor gene）是指一些在细胞繁殖、分化等过程中起负调节作用的基因，丢失或失活时有致癌作用。一般把抑癌基因分为两大类：一类是抑癌基因产物与癌基因产物直接作用；第二类是抑癌基因对癌基因的表达起负调节作用，包括转录和转录后调节。

1. Rb基因产物　其基因产物是105kD的、叫做pRb的核内磷酸化蛋白，在细胞周期的不同阶段，以高磷酸化和低磷酸化两种形式存在。在G1期pBb处于低磷酸化状态；当细胞开始向S期转变时，pRb的磷酸化急剧增加并持续到G2期和M期，之后又回到G1期的低磷酸化状态。pRb可以和细胞转录因子E2F结合，这种复合体参与调节细胞从G1期向S1过渡。大T抗原和E1A、E7蛋白可以同E2F竞争性地与pRb结合，从而阻断细胞从G1期到S期的过渡，因此认为pRb对G0和G1期细胞起负调节作用。

2. p53产物　p53基因编码的p53蛋白是一个53KD的核内转录因子，其作用特点是以四聚体的形式与DNA结合。最近的研究表明，p53对细胞周期发生影响。2个p53等位基因在DNA受损的细胞接受离子辐射或某些药物后会发生突变。这些p53基因发生突变的细胞会进入S期，而含野生型p53基因的细胞在接受上述刺激后，p53表达量升高，细胞周期停止在G1期，从而使细胞有足够的时间进行受损DNA的修复。如果受损的DNA不能被修复，细胞就会进入程序死亡。由此可见，野生型p53发挥着调节细胞周期的功能（见第八章）。

3. wnt－1基因产物　该基因编码一个由345个氨基酸组成的DNA结合因子。蛋白质激酶和癌基因的产物EGR－1接受血清因子刺激可迅速激活细胞增殖。若wnt－1蛋白与特定的DNA序列结合后，就阻止了EGR－1与DNA结合，从而抑制了EGR－1对转录的激活作用，wnt－1蛋白即是以此对细胞生长进行抑制的。

4. dcc基因产物　该基因编码一个分子量为190kD的跨膜磷酸化蛋白，DCC蛋白是细胞浆中的信号传递受体，已经发现多发性息肉结肠癌与dcc基因缺失有关，这种缺失可能使细胞获得生长优势。DCC蛋白与一种细胞粘附因子的结构非常相似，细胞间正常粘附作用的丢失可导致肿瘤细胞的扩散和转移。DCC蛋白可能在恢复细胞的粘附作用上发挥作用。

5. mcc基因产物　mcc基因编码一条829个氨基酸的蛋白质，部分序列与G蛋白结合的一种受体相似。在散发性直肠癌中可检出mcc基因的错义突变。

MCC蛋白可能代替G蛋白受体与G蛋白结合来改变或阻止信号的传送。

6. erb基因产物 在一种禽类成细胞增多症病毒株AEV－ES4/R中，发现了2种病毒基因：一种叫V－erbB；另一种为V－erbA。V－erbB和V－erbA有相反的功能。V－erbB可以促进细胞转化，而V－erbA则阻止细胞进入终末分化，使其停留在高度增殖的未成熟状态。有人把V－erbB归为癌基因类，把V－erbA归于抑癌基因类。后来又证实V－erbA是一种高亲和性甲状腺受体基因。erbA蛋白可与DNA结合，羧基端可与激素结合。甲状腺素正是通过与V－erbA蛋白结合对转录造成影响而对体内的糖代谢、脂代谢以及中枢神经系统发育产生作用，但确切的作用机制尚不清楚。

（五）肿瘤转移调控基因

癌细胞转移是造成癌症病人死亡的主要原因，它是由一系列复杂步骤构成的复杂生物学过程。研究证明，肿瘤的异质性是一普遍现象，即不是所有肿瘤细胞都具有转移能力，转移灶是肿瘤与宿主间一系列复杂作用的结果。已发现ras和myc基因家族和突变型p53基因的异常表达与肿瘤的转移有一定的相关性。msfl基因可能为诱发转移的特定基因。

肿瘤转移过程的形成是受转移相关基因和转移抑制相关基因调控的。凡是能够以某种方式或作用途径抑制肿瘤转移的基因均可命名为转移抑制基因(metastasis suppressor gene)。按照肿瘤转移抑制基因在肿瘤转移过程不同阶段的作用方式，大致可以分成三类：①增强或提高肿瘤细胞在宿主体内的免疫原性；②表达翻译一种直接抑制参与转移过程的某种基因产物；③抑制参与细胞转移连锁反应中的蛋白质或直接作用于转移相关基因的RNA或DNA。

目前，nm23基因被认为是肿瘤转移抑制相关基因的典型代表，不少动物肿瘤模型都显示nm23 RNA在转移性肿瘤细胞中表达下调，在多种人类肿瘤如乳腺癌、肾癌、黑色素细胞瘤、肺癌、结肠癌等的肿瘤组织中，nm23的表达与患者的愈后或肿瘤的淋巴结转移有关。

四、DNA肿瘤病毒与癌

DNA肿瘤病毒是在自然状态下可感染人或动物并能诱发肿瘤、致正常细胞转化的一类DNA病毒。它们分属于5个科（表6－6），结构上皆为20面立体对称型，除疱疹病毒外均无包膜，病毒核心是DNA，外有衣壳蛋白包绕，壳粒数目的多少与病毒的大小有关，如乳头状瘤病毒有72个壳粒、其直径为45～55nm，而疱疹病毒壳粒是162个，其直径可达120～200nm。

DNA病毒感染细胞后可导致细胞死亡，并释放出大量子代病毒，这种被

感染的细胞不发生转化，宿主亦不发生肿瘤。若 DNA 病毒感染细胞后仅将病毒 DNA 整合到宿主细胞 DNA 中而不复制病毒，则可诱发细胞转化或发生肿瘤。DNA 肿瘤病毒的转化基因有如下共同特点：① DNA 肿瘤病毒基因组有早期复制和晚期复制 2 个基因区域，早期复制区与转录有关，晚期复制区主要参与病毒的复制。目前已知，所有 DNA 肿瘤病毒的转化基因均位于病毒基因组的早期复制基因区，如 SV40 和小鼠多瘤病毒转化基因—大 T 抗原基因以及腺病毒转化基因—EIA 和 EIB 均如此。② DNA 肿瘤病毒转化基因均为核内癌基因（nuclear oncogene）其编码产物均为核内致癌蛋白（nuclear onco－protein），如大 T 抗原蛋白、EIA、EIB 蛋白等对细胞均有很强的转化作用。③ 由 DNA 肿瘤病毒转化基因编码的致癌蛋白有相似的氨基酸序列。如腺病毒 EIA 基因编码的 EIA 蛋白有 3 个高度保守区，其中 2 区与大 T 抗原、HPV－16 亚型的 E7 蛋白有很高的同源性，且凡与 EIA2 区同源的蛋白都可能具有转化的功能。

（一）多瘤病毒科与肿瘤

此科包括多瘤病毒、SV40 病毒、BK 病毒和 JC 病毒。如本章前述，SV40 病毒是从培养的猴肾细胞内分离出来的，含 5243 个碱基对，是最小的 DNA 肿瘤病毒，早期（DNA 复制以前）以逆时针方向编码大 T 抗原和小 t 抗原，二者与宿主细胞转化有直接关系。大 T 抗原对细胞转化起启动作用，而小 t 抗原对细胞转化起加强作用。晚期（DNA 复制后）基因组以顺时针方向编码结构蛋白，即 VP1、VP2、VP3 三种衣壳蛋白，它们与细胞转化无关。小鼠多瘤病毒基因组与 SV40 病毒相似，早期编码的大 T 抗原、中 T 抗原、小 T 抗原与细胞转化有关、晚期编码 VP1、VP2、VP3 结构蛋白。BK 病毒和 JC 病毒可诱发新生鼠肿瘤，且体外能诱导多种细胞转化。

多瘤病毒潜伏于野生小鼠中，幼鼠由于接触成年鼠的唾啐或尿液而被感染，产生多个部位如腮腺、颌下腺、乳腺、肾脏和消化道肿瘤。SV40 在体外可使培养的大鼠、仓鼠等细胞发生转化，将 SV40 注射入新生仓鼠可诱发肉瘤。BK 病毒在体外能使鼠类、家兔和非洲绿猴细胞发生转化，注入新生小鼠、仓鼠能产生肿瘤。JC 病毒能使新生仓鼠发生神经胶质瘤。

（二）乳头瘤病毒科与肿瘤

不同乳头状瘤病毒（papilloma virus）形态、结构相似，其直径为 55nm，20 面立体对称，含 72 个壳粒，基因组是双链共价闭合的环状 DNA，大小约 7.9kb。人类乳头状瘤病毒（human papilloma virus，HPV）有 60 多个亚型，不同亚型能诱发人类不同病变，可引起良性和恶性肿瘤。良性肿瘤有乳头状瘤、纤维瘤和纤维乳头状瘤，还有疣；恶性肿瘤有子宫颈癌、舌癌、喉癌等。

HPV－6、HPV－11 与性传递的尖锐湿疣和喉头状瘤有关，HPV－16、HPV－18 与子宫颈癌密切相关，自感染 HPV 至发生原位癌约需 35 年。突变的 HPV 只出现于不全角化的上皮细胞内，表明病毒的成熟与细胞的角化程度有关。

（三）腺病毒科与肿瘤

腺病毒（Ad）一般直径为 70～90nm，20 面立体对称，衣壳由 252 个壳粒组成，其 DNA 为双链线型。依腺病毒同源性、致瘤性及 DNA 中 G＋C 含量不同，将其分为 A、B、C、D、E 5 个亚属。共有 90 多个型别，各亚属对动物致瘤作用不同。将 A 组病毒亚属（A12、18、31）注射给新生仓鼠，4 个月内便可诱发肿瘤。B 组病毒亚属（A3、7、11、14、16、21）除 A11 外，均可使少部分动物在 4～18 个月内产生肿瘤。C、D、E3 组亚属均不诱发动物肿瘤。病毒基因组中 G＋C 含量与该病毒致瘤作用强度成反比。目前已证实，所有人的腺病毒都能在体外转化大鼠细胞，且发现该病毒诱导细胞转化不需全部基因。对腺病毒基因结构和表达产物研究发现，A2 型早期转录区有 E1A、E1B、E2A、E2B、E3、E4 6 个独立区。其中 E1A 和 E1B 与腺病毒其他基因和细胞转化有关，EIA 编码 2 种磷酸化的 T 抗原，为细胞转化所必需。腺病毒虽在自然界分布广泛，且很多亚型对动物具有强致癌性，但迄今尚未肯定与人类肿瘤发生有确切的关系。

（四）疱疹病毒科与肿瘤

1．*单纯疱疹病毒* 这种病毒直径为 120～150nm，球形，衣壳由 162 个壳粒组成，衣壳外覆盖有一层起源于细胞膜、且含有脂质的包膜。该病毒基因组由 175kb 的线性双链 DNA 组成，G＋C 含量较高，分为单纯疱疹病毒 I 型（HSV－I）和单纯疱疹病毒Ⅱ型（HSV－Ⅱ）。两型之间可产生交叉免疫。人类是单纯疱疹病毒的自然宿主，除常见的唇疱疹及目前日益增多的生殖器疱疹外，它与人类宫颈癌的发生关系也越来越受到重视。

2．*EB 病毒*（Epsteia－Barr virus） 是由 Epstein 和 Barr（1964）从非洲 Burkitt 淋巴瘤细胞中分离出的一种病毒，并由此而命名。已证实，EB 病毒与传染性单核细胞增多症、Barkitt 淋巴癌、多发性 B 细胞瘤及鼻咽癌等发生有密切关系。EB 病毒直径为 180nm，含双链 DNA、呈线状，但在细胞染色体外游离的 DNA 呈一环状，EB 病毒在体外能使淋巴细胞长期生长，且可使来源于人或猴的 B 淋巴细胞永生化。

（1）EB 病毒与 Burkitt 淋巴病 英国人 Burkitt 于 20 世纪 50 年代后期发现在非洲儿童中流行一种主要侵犯下颌部的淋巴瘤，称为 Burkitt 肉瘤（BL），Epstein 及 Barr 用电镜发现了瘤细胞中的病毒，后人称为 EB 病毒。现在非洲以外地区包括中国在内已有这种病例的报道。对大量非洲儿童作血清学流行病

学调查，发现在发生 BL 前数年（至少 2 年），患者体内已出现高浓度的癌蛋白（VCA）抗体。在经过培养的 BL 瘤细胞内含有 EB 病毒基因组，除少数病毒 DNA 分子以线性共价整合到宿主细胞染色体以外，大多数 EB 病毒 DNA 以共价闭合环形式游离于染色体之外。BL 的发生是一个复杂的多步骤过程，主要与下述 3 个方面有关：①儿童感染 EB 病毒，使部分 B 淋巴细胞具有 EB 病毒标记；②机体在某种抗原（如月原虫抗原等）刺激下，使 B 淋巴细胞不断增殖；③BL 细胞染色体易位，最常见的是含有 c－myc 基因的第 8 号与 14 号染色体易位，其次是与 2 号或 22 号染色体易位。这几号染色体都携带有编码免疫球蛋白重链或轻链的基因，因易位而使 c－myc 基因也易位到 Ig 基因的临近部位，并受 Ig 基因中增强子的作用而被激活和表达，从而导致 B 淋巴细胞的恶变。此外还可能带有以下机制：染色体易位使 c－myc 基因的转录逃避了抑制因子的作用；易位的 c－myc 也可能处于 Ig 基因以外、尚不清楚的增强子的控制之下。但就目前而言，EB 病毒与 BL 发生过程中的确切作用机制尚待进一步证实。

（2）EB 病毒与鼻咽癌（Hasopharyngeal carcinoma，NPC）　NPC 主要发生于东南亚国家和我国华南地区，其发病规律是幼年感染病毒，20 岁后发病，男性多于女性。血清学研究表明，NPC 病人血清中的 IgG/EA 和 IgA/VCA 抗体的阳性率明显高于正常人，患头颈部其他肿瘤患者和患其他部位恶性肿瘤病人的 IgA/VCA 升高更具特异性。这种抗体出现在 NPC 临床发病以前，其浓度随病人存活时间延长而下降、随肿瘤复发或转移而升高。高分化 NPC 细胞株及低分化 NPC 细胞株都有 EB 病毒 DNA 存在。说明 EB 病毒在 NPC 发生中起重要的病因作用。EB 病毒可能是感染上皮细胞，特别是柱状上皮细胞，并整合到细胞 DNA 中，在某种情况下，细胞发生转化并形成鼻咽癌。但研究资料表明，EB 病毒并不是 NPC 唯一的病因，而是多种因素作用的结果，如环境因素、遗传因素等。

（五）嗜肝病毒科与肿瘤

嗜肝 DNA 病毒科（Hepdnaviridae）包括有人乙型肝炎病毒（HBV）、土拨鼠肝炎病毒（WHV）、地松鼠肝炎病毒（GSHV）和鸭乙型肝炎病毒等。这些病毒有相似的形态和基因组，对肝有特殊的亲和性和致癌性，但都不能在体外转化培养细胞。HBV 有独特的形态，在被感染的细胞内，可见到如下三种形态：小球型颗粒、管型颗粒和大球形内含核酸的颗粒即 Dane 颗粒，前两种颗粒因内无核酸故无传染性。HBV DNA 在肝细胞内以游离或整合于细胞染色体上的形式存在。游离 DNA 表示病毒正在复制，多出现于急性或慢性活动性肝炎患者。在慢性肝炎，尤其是原发性肝细胞癌中，多见整合的 HBV DNA。现

已证实，HBV 长链上有 S、C、P 和 X 4 个可读框，它们分别编码不同的蛋白质。S 区（包括 S1、前 S2 和 S 基因）编码 HBSAg，其成分主要为蛋白多糖。C 区（含前 C 区）编码核心抗原（HBcAg）和 e 抗原（HbeAg）；P 区与其他区有重叠可读框，该区编码产物 DNA 多系酶（或逆转录酶）；X 区位于 C 区上游，编码约为 154 个氨基酸的抗原（HbxAg）蛋白。

五、RNA 病毒与肿瘤

不同类型的 RNA 肿瘤病毒有类似的基因组结构，即从 5′端到 3′端依次包括以下 3 个基因（表 6－9）；①主要结构蛋白基因（gag 基因），负责编码结构蛋白，它首先编码前体蛋白，再进一步切割成病毒装配所需的结构蛋白，如小鼠逆转录病毒 gag 基因编码产生的 p10、p12、p15T、p30 等；②聚合酶基因（pol 基因），负责编码依赖 RNA 的 DNA 聚合酶，即逆转录酶；③包膜基因（env 基因）负责编码病毒颗粒外膜的糖蛋白。上述 3 个基因、携带有病毒复制所需的全部遗传信息，只要这 3 个基因完全表达，病毒即可进行复制。有的病毒还含有癌基因（onc），主要编码转化蛋白，引起细胞转化，诱导肿瘤发生。

表 6－9 RNA 肿瘤病毒基因及其编码产物

5′LTR	gag	pol	env	onc	LTR3′
基因产物	主要结构蛋白	聚合酶	包膜糖蛋白	转化蛋白	
禽类	p19、p27、p12、p15	p10、p65 p32	gp85 gp37	pp60scr	
小鼠	p15、p12、p30、p10	p85		p21	
猫	p15、p12、p30、p10			p84	

RNA B 型肿瘤病毒与肿瘤

小鼠乳腺肿瘤病毒（MMTV）属于 B 型病毒，是从小鼠乳腺瘤组织中分离的。在高乳腺癌 A 系雌鼠的乳腺癌组织及同系雄鼠的精囊、睾丸均能发现此类病毒，在低乳腺癌近交系小鼠中也有此病毒，只是需经放射线照射后或化学致癌剂处理后才能检测到。MMTV 常以原病毒的形式通过生殖细胞（卵子和精子）传递。也可通过乳汁进行水平传播。

乳腺癌是人类妇女的主要肿瘤，Moore 等人报告在人的乳汁内也发现有类

似 MMTC B 型病毒颗粒，内有 RNA 及逆转录酶，以后陆续有人报道了类似的结果，但还不足以证明 RNA B 型肿瘤病毒是人类乳腺癌的病因。现认为遗传因素、激素（主要是雌性激素）平衡失调对引发乳腺癌起重要作用。

人类免疫缺陷病毒（human immunodeficiency virus，HIV）是艾滋病的病原。HIV 的靶细胞是 CD_4 细胞（TH 细胞），由于此细胞受损，导致机体免疫功能破坏，由此而产生一系列机会性感染和恶性肿瘤或其他威协生命的临床表现。

HIV 为球状，含 20 面立体对称结构，直径约 100nm，外膜上有 gp120 和 gp41 两种突起，核心含有两个分子单股正链 RNA、分 HIV－I 型和 HIV Ⅱ，世界范围内流行的主要由 HIV－I 型所致。其基因结构分为结构基因，调节基因及附加基因（图 6－13）。

结构基因含 gag 基因（群抗原基因）、聚合酶基因（pol 基因）、env（包膜基因）等，三者编码病毒的基本结构，gag 基因编码的是一个多聚蛋白，前体 N 端附近有十四烷基化合物，与 N 端分解产物 p17 有关。p17 位于病毒囊内表面。另一个裂解产物是 p24，构成病毒核心的主要结构蛋白，大量存在于病毒颗粒中。p9、p7 也是 gag 多聚蛋白较小的核心蛋白产物，由一前体 p15 迅速裂解而来。pol 基因产物与各种酶的功能相关，如逆转录酶、RNA 酶 H、整合酶或核酸内切酶等。env 基因也编码一种多聚蛋白，即 gp160，随后裂解为 gp120 及 gp41（后者因其功能又称透膜蛋白）。gp120 与细胞 CD_4 分子具有很强的亲和力。一半以上的 gp120 分子是糖类。gp120 及 gp41 分子具有高度的免疫原性，在血清学诊断中极为重要，是 HIV 主要抗原，也可能成为有效的疫苗抗原片段。

调节基因有 3 个即 tat、rev 及 nef，参与病毒复制的调节。这 3 个基因的表现依病毒复制时期及速率不同而异。①tat 基因作用的应答区在原病毒的 LTR，缺乏 tat 基因功能的 HIV 不能复制。tat 蛋白（p14）保持病毒 mRNA 的稳定性。②rev 基因能产生大量病毒结构蛋白所需的信息，能增加 gag 与 rev 的表达，是病毒复制的基础。③nef 是第 3 个主要调节基因，为负性转录调节因子，能抑制所有人类免疫缺陷病毒基因的表达。nef 产物 p27 是一种十四烷基蛋白，HIV 感染者中有一半产生 nef 蛋白抗体。

附加基因存在于 HIV－I 型及Ⅱ中，但这种基因不是 HIV 复制所必须的。附加基因可能在宿主方面起重要作用，其中① vif 在因的作用不明，但已知其编码分子量为 2.3×10^4 ②vpr 不是病毒复制所必需，可编码一大约为 1.2×10^4 的蛋白，存在于病毒颗粒及感染细胞内。③vpu 该基因产生分子量为 1.6×10^4 产物，与细胞膜出芽位点有关。

已知前述，HIV 的复制周期非常复杂，宿主细胞上的 CD_4 抗原与 HIV 包膜蛋白（gp120）有强大的亲和力，具有 CD_4 抗原的细胞称为 HIV－I 的靶细胞，包括 CD_4^+ T 细胞、10%～20%大单核细胞和巨噬细胞、5%～10%的 B 细胞、中枢神经系统的交质细胞、神经细胞和肺泡巨噬细胞等。HIV－I 型通过 gp120 与上述细胞的 CD_4 蛋白结合后，再通过 gp41 的透膜进入细胞，脱去外膜、核心部分进入细胞内，经过逆转录酶的作用，两条单正链 RNA 被逆转录为两条负链 DNA，后者被转运入核内，各种信息的转录及拼接均发生于胞核内。仅 10%的感染细胞中的 HIV－I 能复制，其余 90%的感染细胞则处于病毒的潜伏感染状态中。

第五节　病毒感染与宿主干扰素的合成

1957 年，Isaacs 和 Lindenman 在用鸡胚培养经高温灭活的流感病毒研究该病毒对同源感染性病毒的干扰作用时，发现了一种能抑制病毒复制的可溶性物质。他们将该物质取名为干扰素（interferon）。从此，干扰素在病毒学界和生物学界的很多领域受到广泛重视，并对其进行了逐步深入的研究。关于干扰素基本理论的研究，诸如干扰素的产生机制、干扰素类型、干扰素的作用机理、干扰素的提纯等日益取得重大进展，然而，对干扰素的临床实际意义的评价却一直众说纷纭。1967 年，Merck 研究所的科学家发现人工合成的双链核糖核酸（如 poiy I：C）是一种良好的干扰素诱导剂，对干扰素的研究起到推动作用，给人们带来了利用内源性干扰素解决病毒性疾病防治的希望。尔后干扰素基础理论的研究成果和 Strander 等人关于应用干扰素能有效治疗骨肉瘤的报告，又为干扰素的研究带来了一个新高潮。从此，关于干扰素的研究工作已涉及到多种生物学科：病毒学、细胞学、免疫学、分子生物学、基因工程学、生物化学、肿瘤学等。

一、干扰素

干扰素是一类在同种细胞上具有广谱抗病毒活性的蛋白质，其活性的发挥又受细胞基因组的调节和控制，涉及 RNA 和蛋白质的合成。这一定义的含义如下：1. 干扰素是一类蛋白质。2. 干扰素本身并不能灭活病毒，而是通过激活细胞基因来发挥作用的。也就是说干扰素作用于细胞后，使细胞又产生多种其他蛋白质，从而阻断病毒的复制。3. 干扰素具有较广谱的抗病毒活性。

干扰素对蛋白酶敏感；其分子量在 18 000～100 000 之间，比病毒小，经

40 万克离心 4 小时仍无明显沉降。但不能透过半透膜。它在 pH2～10 的范围内都是稳定的，对热相对稳定，在 40℃可稳定几个月。

人干扰素有 3 种，即 α－、β－和 γ－干扰素。α－干扰素（IFN－α）主要是由白细胞产生的，故称之为白细胞干扰素（leukocyte interferon），用病毒或其他诱导剂可诱导出这种干扰素。β－干扰素（IFN－β）主要是在成纤维细胞中诱导出来，故称之为成纤维细胞干扰素（fibroblast interferon）。γ－干扰素（IFN－γ）以前称之为免疫干扰素 II 型干扰素，可以用促细胞分裂原（mitogens）刺激未致敏的淋巴样细胞而产生，也可用特异性抗原刺激致敏淋巴细胞而产生。IFN－α 和 IFN－β 不仅抗原性不同，而且它们的靶细胞特性亦有不同。用 SDS/PAGE 技术进行电泳分析，发现 IFN－β 有一个活性峰，分子量约为 20 000；而 IFN－α 则是由两种分子量明显不同的成分所组成，分别为 21 000～22 000 和 15 000～18 000。淋巴母细胞既可合成 IFN－α，又可合成 IFN－β。该细胞合成的干扰素称为淋巴母细胞样干扰素（lymphoblastoid interferon），也属于 I 型干扰素。IFN－β 和 IFN－γ 是糖蛋白，但为寡糖侧链。IFN－α是没有糖蛋白侧链的干扰素。由于 IFN－α 和 IFN－β 在 pH2 的强酸环境中稳定，并且在含 SDS 的溶液中不减少活性，为提纯带来了方便。这三种干扰素的免疫学特性区别很大，所以无明显交叉反应。

绝大多数（可能是全部）脊椎动物都可产生干扰素。干扰素能抑制病毒增殖，并且这种作用是非特异性的，即任何一种病毒刺激宿主细胞产生的干扰素均具有广谱的抗病毒作用，既可抑制 DNA 病毒的增殖，又可抑制 RNA 病毒的增殖。然而干扰素的作用却具有相对种属特异性，即干扰素活性在同种细胞中高，在异种细胞中低。这种种属特异性可能与细胞表面受体特异性反应有关。但也有例外，如人干扰素保护牛细胞却比牛干扰素强得多。所有的干扰素既具有抗病毒活性，又具有抗细胞活性；所有的干扰素都具有干扰病毒增殖的作用和调节许多细胞功能的作用。在不同类型的细胞中，干扰素表达活性的程度不同。IFN－α 和 IFN－β 抑制 VSV 和 EMC 增殖的作用比 IFN－γ 强，但是 IFN－γ 抑制呼肠孤病毒和痘病毒增殖的能力比其他二者强。

除了以上三种干扰素外，20 世纪 80 年代又报道了几种干扰素，包括人免疫淋巴细胞产生的一种对酸不稳定的 IFN－α 干扰素以及用植物血凝素（phytohemagglutinin，PHA）诱导人单核细胞所产生的另一种干扰素。这些新干扰素具有某些 IFN－α 的主要特征，但是仅仅在第 21 号染色体是 3 体的人成纤维细胞中才有抗病毒活性。同时，这些新干扰素在抗原性、生物学特性和理化特性上与前三种干扰素不同。在感染病毒的羊膜（amniotic membranes）上见不到任何能被抗 IFN－α 和抗 IFN－β 抗体中和的干扰素。这类干扰素在胚胎

发育期间有着重要作用，它们可能与预防母体对胎儿的免疫排斥有关。

由上可见，干扰素是一个复杂的生物学系统。它的产生虽然在最初是通过病毒对细胞的诱导所导致的，但是，实际上诱导干扰素产生的物质极其复杂。除了病毒外，还有一些细菌、立克次氏体、原虫、高等植物的提取物以及人工合成的多聚物等非病毒物质均可诱导细胞产生干扰素。

干扰素是由诱导物诱导细胞产生的一类细胞蛋白，同时它又是诱导细胞产生另一类所谓抗病毒蛋白的诱导剂。

二、干扰素基因

人基因组中含有许多个 IFN－α 干扰素基因，其中已有 14 个干扰素基因得到证实，并且大多数作了序列分析。此外还有 7 个干扰素假基因也得到证实。

现在知道人的第 2 号染色体上载有IFN－β基因；第 5 号染色体上有IFN－α基因；第 12 号染色体上有 IFN－ γ 基因；第 16 号染色体上有干扰素产量调节基因；第 18 号染色体上有 IFN－γ 基因及其受体基因；第 21 号染色体上既有 IFN－α 基因和受体基因，又有 IFN－β 基因和受体基因。第 9 号染色体的短臂上既有白细胞干扰素基因又有成纤维细胞干扰素基因。干扰素诱导而产生的所谓抗病毒蛋白的基因定位在 21 号染色体长臂上。有些干扰素基因是一前一后纵列排列着，彼此密切靠近。9 号染色体上的一个由 9 937 个碱基对组成的部分含有 2 个完整的 IFN－α 基因（L 基因和 J 基因），这两个基因中间有 4 363 个碱基对的基因区间的间隔序列。L 基因和 J 基因约有 96% 同源性，并且每一个基因的侧边为 3 500 个高度同源的核苷酸序列。所有 α－干扰素基因都没有内含子。

对 IFN－β 基因的了解不如 IFN－α 基因清楚。在第 9 号染色体的短臂中也有一个 IFN－β 基因，该基因也没有内含子。至于在该染色体上是否还有其他的 IFN－β 基因尚不清楚。Sehgal 等人（1983）发现人基因组中还有 6 个 IFN－β 基因，很可能在第 2、5 和 9 号染色体上。在人的 12 号染色体上有一个 IFN－γ 基因，该基因有 3 个内含子。

不同的 IFN－α 基因之间有 80%～90% 的序列同源。IFN－α 和 IFN－β 基因的序列彼此有 30% 相同。IFN－γ 基因与其他 2 种干扰素基因不同源，但是，它却编码与 α－和 β－干扰素相同大小的蛋白。所有干扰素基因编码 166 个氨基酸的蛋白，该蛋白 N－末端有 20～23 个氨基酸的单一序列，该序列是分泌干扰素所必要部分。但是，一般的干扰素分子中没有这种单一序列。143

个氨基酸长的 IFN－α，145 个氨基酸长的 IFN－β，146 个氨基酸长的 IFN－γ是具有活性的干扰素蛋白（表 6－10）。

有几个 IFN－α基因在 5′端非编码区有 75% 的序列同源。而 IFN－β基因的相应区则差别很大。有证据表明该区控制着干扰素基因的表达。

用克隆的人干扰素基因作探针去检查其他动物的干扰素，发现各种哺乳动物类都含有全部的 IFN－α基因，但是却只含有 1 个或 2 个 IFN－β基因。这种情况只有牛科动物（如牛、羊等）例外。非哺乳脊椎动物则含有 IFN－β基因而无 IFN－α基因。

人和动物的干扰素均是一种分泌性蛋白，其基因结构与其他分泌性蛋白的基因有类似之处，即由 5′端非编码区、分泌信号多肽编码区，干扰素多肽编码区和 3′端非编码区组成。人 IFN－α、β5′端非编码区一般有几个保守区，如与转录有关的 TATA 序列、基因转录的控制区或识别区 Benoist－区（GGC/TCAATCT）、IFN－β的－59～－53 CTCTGAA 保守区等。不同型的干扰素基因所编码的信号多肽的氨基酸数目是不同的，α－干扰素为 23 个，β－干扰素为 21 个；γ－干扰素为 23 个氨基酸。而且信号多肽的切割位点也不一样，α－和 γ－干扰素均在 G－C 之间，而 β－干扰素则在 S－M 之间切开。此外，氨基酸数目、分子量、糖基结合位点数在不同型的干扰素中略有差异（表6－10）。

表 6－10　人干扰素基因结构及其编码多肽组成情况的比较

特　点	人 α－干扰素								β 干扰素	γ 干扰素
	A	B	B	D	E	F	G	H		
分泌信号氨基酸数	23	23	23	23	23	23	23	23	21	23
起点氨基酸	C	C	C	C	C	C	C	C	M	C
终止密码	TGA	TGA	TGA	TAA	TGA	TGA	TGA	TGA	TGA	TAA
氨基酸数	165	166	166	166	166	166	166	166	166	143
分子量	19219	19472	19384	19392	19308	19719	19683	19497	20004	17116
内含子	0	0	0	0	0	0	0	0	0	3
S－X－T	0	0	0	0	0	2	0	0	1	2
信号肽切割位点	G－C	G－C	G－C	G－C	G－C	G－C	G－C	G－C	S－M	G－C

三、干扰素合成

干扰素按其来源可分为内源性干扰素和外源性干扰素。靠用诱导物在机体内诱导产生的干扰素称为内源性干扰素。靠投予机体的干扰素称为外源性干扰素基因。目前认为最为有效的干扰素诱导剂是一种人工合成的双链核糖核酸，它由多聚肌苷酸和多聚胞嘧啶核苷酸构成，称为聚 IC（poly IC）。

一般说来，大多数正常细胞不能合成干扰素基因，其原因是这些细胞的基因不能进行转录。但是，某些细胞（如淋巴样细胞）却能够自发地转录干扰素，其原因可能是这些细胞能够自发地合成某些诱导物，或者是它们能进行干扰素基因调控区重排。这些细胞有些产生 IFN－γ，但是，多数是以每毫升 100～1000 单位的量产生 IFN－α。这些自发产生的干扰素是同源产物，即是由单一干扰素基因所产生的。这些淋巴样细胞也可通过诱导物的诱导而产生高水平的干扰素和异源干扰素以及 IFN－β。研究表明自发产生干扰素的细胞多数是 T 淋巴细胞，而 B 淋巴细胞是很少自发产生干扰素的。

在培养基中若有干扰素存在，对细胞产生干扰素是非常有利的。用每毫升含 0.5 单位的干扰素溶液处理细胞便可大大提高干扰素产量，而且这种细胞产生干扰素 mRNA 也早，特别是对 IC 诱导物的作用更是如此。在用引子干扰素处理了的细胞中，干扰素 mRNA 的稳定性并未增加，但是它的 mRNA 能更有效地进行翻译。目前，加引子的办法已广泛用于 IFN－α 和 IFN－β 的大批量生产。

真正有广阔前景的生产干扰素的工艺应该是基因工程干扰素。IFN－α、β、γ 的基因工程研究已取得成功。已经构建了几种嵌合 IFN－α 基因，并且发现其表达产物在不同的哺乳动物细胞内和敏感动物体内抑制不同病毒复制的能力很不相同。

干扰素基因已被克隆到细菌、酵母、昆虫和哺乳动物细胞的表达载体之中。所用启动子有 λpL 启动子、多角体病毒蛋白启动子和 trp 启动子。还有人将干扰素作为融合蛋白与 LacZ 基因一起表达。这种方法不仅表达了克隆的干扰素基因，而且表达了生化合成的干扰素基因。Tanaka 和他的同事们（1983）用生化方法合成了 IFN－γ基因，并将其于 Lac UV5 启动子的控制下，在 *E·coli* 中得到了成功表达。干扰素的产量不宜太高，最好是每立升 10^9 单位。在 *E·coli* 中表达的干扰素，它没有信号链，没有糖基化，具有可靠的生物学活性。

干扰素基因已被克隆到酵母细胞。这种转化的酵母细胞每细胞能合成 10^6 以上的具有生物学活性的干扰素分子。Derynck 等（1983）将几个 IFN－α 基

因和 IFN－γ基因插入酵母质粒，在 3－磷酸甘油酸激酶启动子的控制下，表达出没有信号序列而具有生物学活性的干扰素，产量为每立升约 2500 万单位。

用克隆的干扰素基因转染哺乳动物细胞，表达干扰素的研究获得成功。转染哺乳动物细胞的载体有牛乳头状瘤病毒、Psv2－neo 和一种含有 SV40 和 dhfr 的 pKGR 衍生质粒。用含有 IFN－β基因的一种 36kb 的人 DNA 片段转染小鼠 L 细胞后，该细胞即使没有外源诱导物的作用亦能短暂地表达人干扰素；若以 dsRNA 作诱导物，该细胞则可大大地增加干扰素基因表达效果。与此同时，与干扰素基因邻近的基因也被一起诱导表达，然而这种诱导作用与干扰素的产生无关。感染病毒后或用聚 IC 处理后，克隆的干扰素基因还会大大增加细胞转录作用。

转染到哺乳细胞中的克隆基因产生干扰素的水平不太高。大多数情况下，每天每毫升约为 100 000 单位，这个量比固有干扰素基因所表达的干扰素量少得多。启动子的改进是利用克隆基因合成干扰素的主要进展。在正常情况下，控制干扰素基因的成分只由所谓“破坏性”诱导物激活。病毒感染和聚 IC 通常是杀细胞的，所以，干扰素只能短期合成，通常不超过 48 小时。Haynes 和 Weissman（1983）将含有 dhfr 的重组质粒导入了 CHO 细胞，并且在含有氨基喋呤的培养基中进行筛选，分离到了含有多质粒拷贝的克隆，这些克隆每天毫升能产生 20 000～100 000 单位的干扰素，如此产量能持续数月。因此，构筑持续生产干扰素的细胞系或者是在非“破坏性”诱导物刺激下产生干扰素的细胞系应该是一个大的突破。

近来已经纯化了许多种干扰素。最重要的提纯步骤包括蓝色葡聚糖与 Cu 和 Zn 螯合物层析，一定孔径的玻璃珠吸附层析和既用多克隆抗体又用单克隆抗体进行的抗体亲和层析。使用高性能液相色谱层析和 SDS－聚丙烯酰氨凝胶电泳（SDS－PAGE）进行提纯的程序能同时提纯几种干扰素。

四、干扰素的抗病毒活性

干扰素具有广谱的抗病毒作用，它能诱导机体产生抗病毒状态，其抗病毒作用是通过作用于细胞诱导出所谓的抗病毒蛋白（AVP）而实现的。IFN 与受体结合后（至少存在有两类不同的 IFN 受体），起动酪氨酸激酶等几种不同蛋白质合成，并形成 DNA 结合转录因子，与 IFN 诱导的基因的 5′调控区中 14bp 序列结合，进一步启动各种抗病毒蛋白的合成。抗病毒作用比较清楚的由 IFN 诱生的抗病毒蛋白至少有 3 种，分别为 P1/eIF－2α、2－5A 合成酶及 Mx 蛋白。

P1/eIF－2α是一种激酶，在 IFN 作用过的细胞中的浓度比正常细胞要高

20多倍。P1/eIF－2αdsRNA（RNA病毒的复制中间型）的作用下，使其本身的P1亚单位磷酸化，而后进一步激活全酶，导致宿主细胞蛋白质合成起始因子elF－2的α亚单位磷酸化，使之失活。失活后的eIF－2α不能启始病毒蛋白的合成，因而抑制病毒的复制。

有些病毒可逃逸P1 /eIF－2α的作用。如腺病毒编码一低分子量的RNA与这种激酶结合，可防止dsRNA激活P1/eIF－2α；呼肠孤病毒的σ3衣壳蛋白可牢固地与本身的dsRNA结合，使之不能与P1/eIF－2α结合；某些病毒还可降解（如脊髓灰质炎病毒）、抑制（如HIV、流感病毒、痘苗病毒）P1/eIF－2α的合成。

1974年由kerr等发现在IFN作用过的细胞中，2－5A合成酶的含量比正常细胞中高出50～100倍。2－5A［即（2′－5′）pppA（pA）n的简写］可与内源性RNase L结合。RNase L是哺乳动物细胞质中一潜伏状态的RNA酶，相对分子质量约185 000（185kDa），与2′－5′A结合后被激活。现知（2′－5′）pppA（pA）n中2′自由－OH基团对RNase L的激活起重要作用。激活后的RNase L主要在UA、UG、UU部位切割RNA。因此，2－5A主要降解病毒mRNA，抑制病毒蛋白的翻译。

Holler等在研究流感病毒的致病性时发现，有的种系的小鼠对流感病毒敏感，有的不敏感。经IFN作用后，有的小鼠不感染流感病毒，而有的种系的小鼠仍对流感病敏感。表明在某些种系的小鼠细胞中含有一由IFN诱导表达的基因。以后证实在IFN作用后对流感病毒不敏感的小鼠中存在一Mx基因，IFN作用后可诱生一相对分子质量为74～78kDa的GTP结合蛋白，称之为Mx蛋白。因此，认为IFN在这些小鼠诱生的抗流感病毒状态是通过Mx蛋白来起作用的。用Mx基因转染Mx基因缺陷型小鼠后，该小鼠同样可抵抗流感病毒的感染。而且直接将Mx蛋白加入Mx缺陷型细胞系中，即使没有IFN的作用，该细胞系也可抵抗流感病毒的感染。表明IFN的作用是启动Mx基因的表达。

干扰素的抗病毒活性具有如下特点：（1）干扰素的抗病毒作用是抑制而不是杀灭；（2）有相对的种属特异性；（3）不同细胞对干扰素的敏感性不同；（4）不同病毒对干扰素的敏感性也不相同。

干扰素除了具有抗病毒作用外，还对细胞的功能有明显的影响；与机体的免疫系统、免疫状态有明显的关系；同时有不少文献报告了干扰素具有抗肿瘤的作用。其机制包括如下4个方面：（1）干扰素抑制肿瘤病毒的繁殖；（2）干扰素直接抑制肿瘤细胞的生长；（3）调动机体的免疫系统杀伤肿瘤细胞；（4）抑制癌基因的表达。

小　结

病毒是一种非细胞形态的微生物，其基因组分为DNA和RNA两大类。T偶列噬菌体基因组为双链DNA，具有末端重复序列，分早期、晚期两个区。T奇数噬菌体基因组为双链线状分子，分早、中、晚期三部分。λ噬菌体为线状双链DNA，两端具有12个碱基尾巴，可形成环状，在宿主基因组中具有特异性的整合位点。

在哺乳动物双链DNA病毒基因组中，猴肾病毒（SV40）基因组较小，为环状双链DNA分子，长5243bp，并与组蛋白相连结。腺病毒DNA长36kb，有末端颠倒重复结构，单链通过分子内碱基配对可形成锅柄形状。5′端与蛋白质结合。痘病毒基因组较大，约187kb，一端的两条单链末端通过磷酸二酯键共价结合在一起，变性可形成一个连续的单链环状分子，有较长的末端重复序列。花椰菜花叶病毒基因组长8024bp，具有3个位于负链上的编码区。在单链DNA病毒中，单链噬菌体感染宿主细胞后先合成互补链，形成双链环状DNA分子的复制型分子。

RNA病毒中有单链RNA噬菌体如MS2和Qβ，MS_2基因组较小，为单链RNA，而Qβ为负链RNA，0需转录成mRNA后，才能完成蛋白质合成过程。小RNA病毒的基因组都是正链RNA，脊髓灰质炎病毒有一个编码多蛋白序列，其病毒基因组含9个基因。棒病毒基因组为单负链RNA，如狂犬病毒、水泡性口炎病毒等。大RNA病毒基因组如副粘病毒为负链RNA，流感病毒由8条RNA分子组成，每条都编码不同的蛋白质，且能形成锅柄结构；卫星病毒可能是帮助病毒的一种分子寄生物。双链RNA病毒中，有桑蚕的细胞质多角体病毒，哺乳动物呼肠孤病毒、轮状病毒，其基因组由10～11条双链RNA构成，每条编码不同的蛋白质。逆转录病毒亦为单正链RNA病毒，两端为长末端重复序列，含有qag、pol、env、tax等基因。

它通过逆转录合成DNA，然后再整合到宿主染色体上，随宿主基因组的复制而复制，作为宿主基因组的一部分传给子细胞。线状双链DNA病毒，如腺病毒等复制时不形成双向复制叉，双链打开后能用任一链作模板，按5′→3′方向复制，所以一次只能合成一条单链。痘病毒复制是从基因组内的复制启始区开始，最后通过双向复制而成。环状双链DNA病毒如SV_{40}的复制，首先从惟一的一个复制起点开始，两个复制叉朝复制起点相对的一面双向进行，一条链按5′→3′方向连续进行，另一条则先合成冈崎片段再连接成连续的DNA。复制启始过程需要病毒基因编码的大T抗原或小T抗原。乙型肝炎病毒基因

能通过逆转录进行复制。

所有单链 DNA 噬菌体都采用滚环模型将复制型基因组分解成能包裹到病毒粒子中去的单正链 DNA，或用于进一步复制。细小病毒基因组末端能形成 T 型结构，折叠后又可形成下一轮复制的引物。正单链 RNA 病毒复制时需要 RNA 复制酶，逆转录病毒以 DNA 为模板，在逆转录酶的作用下合成 RNA，再整合到宿主染色体上，成为原病毒，再由 RNA 聚合酶转录和经加工后成为 mRNA，在胞质中装配为病毒粒子。

双链 DNA 噬菌体基因组的表达都是分阶段进行的，T 噬菌体、λ 噬菌体启动子与宿主大肠杆菌基因启动子无同源性；腺病毒 EIA 基因是一个早期转录基因，对转录起调节作用 SV40 早期基因在感染过程中一直表达，而晚期基因只在病毒复制时才表达，增效子对早晚期基因的转录也都具有重要作用。负链 RNA 病毒包括水泡性口炎病毒从 3′开始转录，连续转录数个基因，共用一个启动子，N 蛋白可调节正链 RNA 从转录到复制之间的转变。流感病毒的转录需要帽子结构，而帽子结构是从新合成的宿主细胞 mRNA 获得的。

所有由病毒基因组指导合成的蛋白质都是在宿主细胞质核糖体上合成的，在合成策略上不同病毒有所不同。有的先合成一条大的多聚蛋白分子，再加工成几条短的多肽，有的是合成多条 mRNA 分子，再翻译成不同的病毒多肽，如流感病毒。第三种方式是病毒基因组 RNA 首先转录成若干个单顺反子 mRNA片段，然后再各自进行翻译。在所有单正链 RNA 病毒中，同一条 RNA 分子链可作为蛋白质合成的信使，也可作为基因组物质，但不能同时用作翻译和复制。

肿瘤是一种异质性的细胞异常增生性疾病，有良性肿瘤、恶性肿瘤之分。癌的发生是一个多因素、多基因、多步骤的复杂的生物学过程。癌细胞不断生长、分裂却不发生功能分化；引起肿瘤发生的始动因素包括化学因素、物理因素、生物因素、遗传因素等。

化学致癌物中有烷化剂类、多环芳烃类化合物。芳香胺类化合物、氨基偶氮类染料、亚硝胺类化合物、促癌物与致癌物虽然在致癌过程中都发挥重要的作用，但它们的生物学作用特点各不相同，物理致癌因素主要是指电离辐射和非电离辐射，包括 γ 射线、X 射线、宇宙射线和放射性核元素以及紫外线。病毒主要靠将其基因整合到宿主细胞 DNA 中去改变细胞 DNA 而发挥致癌作用。按照肿瘤病毒所含核酸的不同，分为 DNA 肿瘤病毒和 RNA 肿瘤病毒。RNA 肿瘤病毒又称为逆转录病毒，分为 A、B、C、D 4 种不同的类型。DNA 肿瘤病毒是指一类在自然界能感染动物或人，并在体内能诱发肿瘤，在体外可诱导正常细胞转化，所含核酸为 DNA 的病毒。它分属于 DNA 病毒的多瘤病毒科、

乳头状病毒科、腺病毒科、疱疹病毒科和嗜肝病毒科。

存在于正常细胞中未被激活的癌基因称为原癌基因，激活后的原癌基因即为癌基因。细胞癌基因（c－onc）与病毒癌基因（v－onc）是同源物。目前认为，RNA 致瘤病毒中的 v－onc 是在长期生物演化过程中，病毒从宿主细胞中摄取的，经过加工后成为致瘤病毒的。动物致瘤病毒的研究导致人的 c－onc 的发现，这是病毒肿瘤学的重要贡献。

现已发现了 30 多种 c－onc 产物，按其功能大致可分为：生长因子类、膜受体类、传导分子类、核内蛋白类以及其他类等。c－onc 属于正常的细胞结构基因之一，虽然在正常情况下也能进行表达，与细胞的生长、分化、代谢等有关，但相对说来仍处于抑制状态。在一定条件下被激活之后才能导致肿瘤的发生。抑癌基因是指一些在细胞繁殖、分化等过程中起负调节作用的基因，丢失或灭活时有致癌作用。肿瘤转移过程的形成受转移相关基因和转移抑制相关基因的调控。

干扰素是一类在同种细胞上具有广谱抗病毒活性的蛋白质。人干扰素有 3 种，即 α－干扰素（IRN－α）、β－干扰素（IFN－β）、γ－干扰素（IFN－γ）。人基因组中含有许多个 IFN－α 干扰素基因，其中已有 14 个干扰素基因得到证实，所有的 α 干扰素基因都没有内含子。IFN－β 基因也没有内含子，而 IFN－γ 基因有 3 个内含子。干扰素具有广谱抗病毒作用，它能诱导机体产生抗病毒状态，其抗病毒作用是通过作用于细胞诱导出所谓的抗病毒蛋白（AVP）而实现的。干扰素具有抗肿瘤的作用，其机制包括如下 4 个方面：（1）干扰素抑制肿瘤病毒的繁殖；（2）干扰素直接抑制肿瘤细胞的生长；（3）调动机体的免疫系统杀伤肿瘤细胞；（4）抑制癌基因的表达。

复　习　题

1．比较 λ 噬菌体、人乙型肝炎病毒和逆转录病毒基因组的结构特点，它们在基因组复制、转录和蛋白质合成方面有何异同？

2．单链 DNA 噬菌体主要有哪几种？其基因组复制有什么特点？

3．诱发肿瘤的因素有哪些？DNA 肿瘤病毒致细胞转化有什么特点？

4．从哪些方面可以看出病毒的生命循环必须依赖宿主细胞？是否所有病毒的基因都由病毒自身编码的转录酶进行转录？许多噬菌体基因的启动子与其宿主基因的启子不同，这是否说明这些基因不由宿主细胞的 RNA 聚合酶转录？

5．何谓原癌基因、细胞癌基因？它们与哪些病毒的癌基因有关？

6．干扰素在什么条件下合成？干扰素抗病毒活性和抗肿瘤作用各有什么特点？简述 α、β、γ 干扰素基因的结构特点。

第七章

脊椎动物免疫的基因控制

第一节　脊椎动物的免疫反应

免疫学（immunology）几乎是与医学微生物学和兽医微生物学同时诞生的，20 世纪 60 年代以来，免疫学无论在理论上和实践上都有了飞跃的发展，其研究范围已渗透到整个生命科学领域，从而形成了一门独立、富有生命力的新兴学科。本章在简明介绍免疫学的一般原理的基础上，重点讨论免疫的遗传控制机理。

一、免疫的概念

在 Jenner 和 Pasteur 等人的古典免疫概念中，免疫被看作是对微生物感染的抵抗力和对同种微生物再感染的特异性的防御能力。但很多研究证实，有很多免疫现象与微生物无关，如血型、对异种蛋白的反应、过敏反应、肿瘤免疫等。因此，免疫的概念实际上已大大超过了抵抗感染的范围。

现代免疫学的概念是：免疫是机体识别和清除非自身的大分子物质，从而保持机体内外环境平衡的生理学反应。此种本能是动物长期进化中形成的。

（一）免疫的基本特性

识别自身物质和非自身物质是免疫应答的基础。机体的这种识别十分精密，不仅能识别异种蛋白质，甚至同一种动物的不同个体的组织细胞也能识别，从而出现对异体组织移植的排异反应。只有基因型完全相同的个体，如纯系动物或 MHC 相同的兄弟姐妹间才能进行异体组织器官的移植而不被排斥。像这样微细的差别即使用目前最先进的理化测试手段也是无法区别的。识别功能对机体的健康是十分重要的，识别功能的降低就会延缓免疫应答功能的启动，从而使机体降低或丧失对传染或肿瘤的防御能力。识别功能的紊乱，更易招致严重的生理失调，如把自身的蛋白质或细胞误认作“非己”，就会造成自

身免疫病。

1. 特异性　机体免疫系统不仅能识别自身和非自身，还能识别非自身物质间的微小差异，甚至某些同分异构体，如氨基苯磺酸基位置（邻位、对位、间位），酒石酸的旋光性（左旋、右旋、消旋）等均能予以识别，这就奠定了免疫血清学技术在生物活性物质的定性、定量和定位中的不可替代的地位。

2. 免疫记忆（immunological memory）　抗原进入动物体内以后，经过一段时间的潜伏期，血液中出现抗体，徐徐增加并达到顶峰，随之逐渐下降以至消失。当抗体消失后，用同源抗原加强免疫时，能迅速产生比初次接触抗原时更多的抗体。这一现象表明机体具有免疫记忆能力。细胞免疫也同样具有免疫记忆，这是因为在初次接触抗原之际，除刺激机体形成产生抗体的细胞以外同时也形成了免疫记忆细胞（immune memory cell）。动物被某些传染病感染后或用某些疫苗免疫后之所以能获得长期的免疫力，主要是免疫记忆之功。

（二）免疫的基本功能

1. 抵抗感染　脊椎动物免疫功能正常时，能充分发挥对病原微生物感染的抵抗力，对侵入体内的病原微生物，能迅速激发细胞免疫和体液免疫，大大加强对微生物的清除能力，并予以歼灭。但当免疫反应异常亢进时，反而导致传染性变态反应；而免疫状态低下或免疫不全症患者又往往出现反复感染。

2. 自身稳定　在新陈代谢中，每天都有大量的细胞衰老死亡，这些失去了功能的细胞积累在体内，必然会影响正常细胞的活动。免疫系统的第二个功能就是消除这些细胞，保持机体正常细胞的生理活动，使机体的各部门能精确地执行正常功能，充满活力。当细胞衰老死亡时，可刺激机体产生自身抗体，以及时清除这些细胞。但是，如果此种机能异常亢进，产生过多的自身抗体，从而危及正常细胞时，就会引起自身免疫病。

3. 免疫监视（immune curveillance）　动物最危险的“敌人”来自于身体内部。正常细胞在化学、物理及病毒等致癌因素的诱导下可以变为肿瘤细胞，免疫的第三个功能就是严密监视这些肿瘤细胞的出现，一旦出现就能立即予以识别，并调动免疫系统在其尚未形成肿瘤组织之前将其歼灭，免疫功能降低或受抑制就会使肿瘤细胞大量增加，从而出现临床上的肿瘤。

免疫是一个动态过程，在抗感染免疫中，这个过程包括致病微生物侵入动物机体——被免疫系统识别——淋巴细胞选择性活化——产生特异的体液免疫和细胞免疫——机体进入免疫状态——病原微生物被清除。这是初次免疫应答过程，在这以后，机体的免疫状态仍可保持一定时间，若一定时间内再次接触同类微生物，即可触发再次免疫应答，迅速将侵入者清除，并使机体获得更强的、更长期的免疫力。因此，用疫苗来模拟这一过程，就可使人或动物获得人

工自动免疫，这一过程亦即免疫预防的全过程。

免疫应答通常分体液免疫和细胞免疫，但二者相互协作，相互制约形成一个不可分割的整体，抗体的形成有赖于辅助性T细胞的介导，巨噬细胞和K细胞对靶细胞的杀伤作用，也需要借助于抗体的识别。此外，两者还存在着巧妙的、精密的自动调节网络。动物机体能调节免疫应答水平，以免反应过强或过弱而造成损害。所以，免疫亦有两重性，它既是重要的防御反应，但亦往往陪伴着有害的病理过程，如传染性变态反应是机体为了清除抗原异物的一种特殊形式的免疫反应，但其反应程度往往超越了正常的生理范围，从而导致功能紊乱和组织损伤。

二、免疫系统与免疫反应

由主宰或执行机体免疫功能的器官、组织、细胞和分子所组成的一个系统称为免疫系统（immune system）。免疫系统由免疫器官、免疫细胞和免疫分子所组成，这里主要介绍免疫器官和免疫细胞，免疫分子在后面有关章节中介绍。

（一）免疫器官

1．中枢免疫器官（central immune organ）　中框免疫器官又称一级免疫器官，是免疫器官发生、分化和成熟的场所，对外周免疫器官的发育起主导作用，中枢免疫器官包括骨髓、胸腺、鸟类还有腔上囊。胸腺（thymus）由胚胎第Ⅲ、Ⅳ对咽囊的内胚层分化而来。它位于胸隔上部、胸骨后方，分左右两叶。胸腺的大小和结构随年龄和机体的状态而变化，出生时重量约为10～15g，出生后两年内迅速增大，此期为胸腺活动的高峰期，此后逐渐增大，至青春期最重，约30～40g，青春期以后胸腺开始缓慢退化。进入老年，胸腺组织大部分被脂肪组织所取代，但仍残留一定的功能，胸腺的这种随年龄增长而萎缩的现象，称为生理性胸腺萎缩。

胸腺左右两叶的表面有结缔组织的包膜，伸入胸腺实质，将胸腺实质分成若干小叶，小叶周围为皮质，深部为髓质，相邻的髓质彼此相同。在胸腺内分化成熟的淋巴细胞称为胸腺依赖性淋巴细胞（tympus－dependent lymphocyte），简称T细胞。皮质是T细胞急剧增殖和大量死亡的区域。在胸腺内分化的T细胞只有1%，最终成为成熟的T细胞，并经血流迁移至外周淋巴组织中。T细胞的死亡是一种生理性死亡，死亡的细胞迅速被巨噬细胞吞噬、消化。这种死亡可能对消除自身反应性T细胞克隆起重要作用，它有助于消除禁忌克隆（forbidden clone），维持机体自身稳定以及调节体内各种T细胞亚群

间的合适比例。

胸腺是诱导T细胞分化、成熟的主要器官。前胸腺细胞在胸腺微环境中，在胸腺激素的诱导下发育成熟，并以恒定的比例（1%～2%）迁出胸腺外，定居于外周器官或组织。胸腺激素由胸腺内的网状内皮细胞产生，在T细胞的分化和调节中起重要作用。

骨髓（bone marrow）是人和其他哺乳动物的造血器官，也是各种免疫细胞的发源地，骨髓的多能干细胞能分化为髓样干细胞和淋巴细胞，为重要的免疫器官。人和哺乳动物中重要的免疫细胞除T细胞在发育过程中是从骨髓到胸腺进行驯化以外，其余免疫细胞几乎都是在红骨髓中发生和发育的。法氏囊是鸟类特有的器官，来自于骨髓的干细胞，经血流进入法氏囊后，分化成为囊依赖性淋巴细胞（bursa dependent lymphocyte）。简称B细胞，参与体液免疫。哺乳动物并不存在独立的囊类结构，骨髓的一部分作用相当于法氏囊的功能。

2. 外周免疫器官　外周免疫器官包括脾脏、淋巴结和消化道、呼吸道及泌尿生殖道的淋巴小结。它们是T、B细胞定居和对抗原刺激进行免疫应答的场所，它们能迅速捕获抗原，并为处理后的抗原以最大的机会提供给免疫活性细胞。

淋巴结（lymph node）　呈圆形或豆状，遍布于淋巴循环路径的各个部位，以便捕获从身体外部进入血液—淋巴液的抗原。它由网状组织构成支架，包有结缔组织包膜，其内充满淋巴细胞、巨噬细胞和树突状细胞；其中，B细胞主要位于浅部淋巴小结内及深部髓质的髓索内，淋巴小结周围和副皮质区是T细胞的主要集中区。

脾脏（spleen）　是体内最大的免疫器官，具有造血和免疫双重功能。脾脏是血液通路上的滤过器官，可滤除血液中的抗原异物及自身衰老、死亡细胞的碎片等物质，脾的实质可分为红髓和白髓两部分，红髓由脾索和脾窦组成，脾索呈网状，脾窦是分布于脾索之间的窦腔，血细胞可以从脾索进入脾窦，白髓是由密集的淋巴组织构成，沿着动脉分布，分散于红髓之间。白髓包括动脉周围淋巴组织鞘（淋巴鞘）和淋巴小结（脾小结），前者是T细胞集聚地，后者主要由B细胞构成，由外界进入血液中的抗原大部分在脾脏中被截获。脾脏中B细胞占60%，T细胞占40%，所以脾脏是产生抗体的主要基地。

扁桃体（tonsil）　是一种重要的外周免疫组织，其表面被有复层扁平上皮，上皮向内凹陷形成许多隐窝，隐窝周围有许多淋巴小结和弥散淋巴组织，扁桃体是最容易受病原异物侵害的部位。

肠道相关淋巴样组织　是存在于消化道淋巴组织的总称，包括集合淋巴结和阑尾。此外，禽类还有哈德尔氏腺，它是禽类眼窝内分泌性腺的组织之一，

亦称副泪腺，能接受抗原物质的刺激，独立地完成免疫应答反应，在眼部和上呼吸道局部免疫中有重要意义。

（二）免疫细胞　参与免疫应答的细胞统称为免疫细胞，T 细胞、B 细胞、N 细胞、K 细胞、单核巨噬细胞、各种粒细胞等。T 细胞和 B 细胞均为小淋巴细胞，受抗原刺激后细胞变大，进一步分化、增殖为淋巴母细胞，B 细胞分化为浆母细胞，直至形成浆细胞，产生抗体参与体液免疫。T 细胞经抗原刺激后，进一步分化为不同的亚类，包括细胞毒性 T 细胞（cytotoxic T lymphocyte，CTL 或 Tc）或杀伤性 T 细胞（killer T cell，Tk），抑制性 T 细胞（suppressor T lymphocytes，STL 或 Ts），辅助性 T 细胞（helper T lymphocyte，TH）。

三、抗原及抗原性

1. *抗原的概念*　凡是能刺激机体产生抗体和致敏淋巴细胞、并能与之结合从而发生特异性免疫反应的物质称为抗原（antigen），刺激机体产生抗体和致敏淋巴细胞的特性称为免疫原性（immunogenicity）；与相应抗体结合发生反应的特异性称为反应原性（reactivicity）或免疫反应性（immunoreactivity）。某一蛋白分子具有抗原性是指它同时具有免疫原性和反应原性。这种抗原称为完全抗原，只有反应原性而没有免疫原性称为不完全抗原，或称半抗原。

2. *构成抗原的条件*　构成抗原的物质，必须具有异源性，分子量常大于 10 000 以上，在分子构型上具有复杂性等。

抗原通常是非自身物质。动物蛋白质的免疫原性一般与动物在种系分类上的亲疏程序呈负相关，血缘关系越远，生物种系差异越大，则抗原性就越强。例如鸭的蛋白质对鸡是较弱的免疫原，而对兔则是良好的免疫原。同种异体之间有时也有免疫原性，如血型抗原，主组织相容性复合体等。因此，在不同个体之间进行组织和器官移植时，可引起移植排斥反应。自身组织通常没有免疫原性，只有一些在正常状态下不与免疫系统接触的自身隐蔽物质如眼球晶体蛋白、精子蛋白和甲状腺蛋白等，一旦因受伤、感染等原因进入血液时，亦能引起机体的免疫反应。

抗原的免疫性与分子的大小密切相关，常大于 10 000 以上，分子量小于 5 000者一般免疫原性很弱，分子量在 1 000 以下的小分子为半抗原，本身通常没有免疫原性，但当它们与蛋白质载体结合后即可获得免疫原性。

抗原物质除了要求有一定的分子量以外，还要求必须有一定的分子结构和空间构象。分子结构和空间构象复杂者抗原性强，例如明胶分子量虽然很大

(高达 100kD)，但结构上主要是直链氨基酸，稳定性差，在体内易被分解为小分子物质，故抗原性很弱。抗原分子的空间构象，即基团之间的相互立体关系是决定抗原性最重要的因素。

3. 抗原特异性与抗原决定族

抗原特异性　抗原所引起的免疫应答具有极强的针对性，一种抗原物质刺激机体产生相应的抗体，这种抗原只能与相应的抗体相结合发生反应，此即抗原的特异性。这种特异性取决于分子表面具有特殊立体构型的、有免疫活性的化学基团，这些基团易为机体的免疫应答所针对，是与抗体相结合的部位，这个区域或基团称为抗原决定簇 (antigenic determinant)，又称为抗原表位 (epitopes)，抗原特异性决定于抗原决定簇的性质、氨基酸或碳水化合物的种类、序列及空间立体结构。每个决定簇由 4～7 个氨基酸基团组成，多糖抗原则由 5～6 个已糖残基组成。一个抗原分子上可能有多个抗原决定簇，若多个决定簇抗原性一致，则为单特异性决定簇，若不一致，则为多特异性决定簇。暴露于分子表面的决定簇，能与免疫活性细胞相接近，对激发机体的免疫应答有决定意义；而在分子内部的决定簇，只有在用酶轻度消化后才能暴露，如牛血清蛋白，分子量为 69 000，有 18 个决定簇，但只有 6 个暴露在外面。

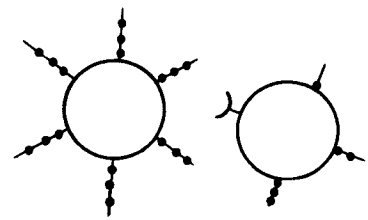


图 7－1　抗原的单特异性和多特异性决定簇

4. 抗原类型　天然抗原种类繁多，有数百万种之多，但就其化学性质来说，主要有以下几类 (表 7－1)。

表 7－1　天然抗原的化学分类

类　别	天　然　抗　原
蛋白质	血清蛋白 (白蛋白、球蛋白) 酶、细菌外毒素、病毒蛋白
脂蛋白	血清 α、β－脂蛋白
糖蛋白	血型物质、组织相容性抗原
脂　质	结核菌的磷脂质和糖脂质
多　糖	肺炎球菌的荚膜
旨多糖	革兰氏阴性细菌的细胞壁 (O 抗原、内毒素)
核　酸	核蛋白等

人工抗原　是指人工构建的抗原，它包括两类，一类是结合抗原，即半抗

原与载体结合后所形成的抗原，通过化学方法将某些不具备免疫原性的半抗原结合到某种蛋白载体上，构成了新的抗原性物质，结合上的半抗原即成为该抗原分子上的一个抗原决定簇。由于这一技术的发展，使得几乎所有对生命活动有重要意义的小分子活性物质均可以用此法制备抗血清，从而利用极为敏感的免疫学方法对此类物质进行超微量分析。

合成抗原　用化学方法将某些已知氨基酸按一定序列聚合成大分子多肽，使其具有免疫原性，可以使用一种氨基酸，也可使用多种氨基酸，可以是直链的，也可是带有侧链的。合成抗原一方面可以用于抗原结构、抗原特异性等免疫理论的研究，同时合成肽所构成的第三代疫苗可能为将来的人工自动免疫提供更为优秀的无副作用的生物制品。

在天然抗原中，微生物抗原是最重要的一类抗原，如细菌抗原、病毒抗原、毒素抗原等。

第二节　B细胞免疫反应

一、抗体、单克隆抗体及其基因工程抗体

（一）抗体的结构与功能

抗体（antibody）　是在抗原和免疫系统相互作用下，由B细胞产生的能与相应抗原发生特异结合的免疫球蛋白（Immunoglobulin，Ig），通常存在于血液等多种体液中及某些细胞的表面上。抗体最显著的特点是具有特异性和多样性。根据重链稳定区的分子结构和抗原特异性不同，人类Ig可分为五类，即IgG、IgA、IgM、IgD、IgE。血清中以IgG含量最多，IgA、IgM较少，IgD、IgE仅微量。

抗体分子的基本结构是四条肽链，其中两条为相同的重链（heavy chain，H链），其分子量较大，由440个氨基酸组成，另两条为相同的轻链（light chain，L链），其分子量较小，只含有214个氨基酸。重链上结合有不同量的糖，故Ig属糖蛋白。轻链根据其C区的化学结构和抗原性差异分为κ（Kappa）和λ（Lambda）二型，一个Ig分子的二条L链不是κ型就是λ型，重链亦根据C区的分子结构和抗原性差异分为5类，即α、δ、ε、γ、μ5种，据此将Ig分为IgG、IgA、IgM、IgD、IgE。每条肽链分氨基端（N端）和羧基端（C端），四条多肽链靠二硫键连结呈丫字型（图7-2）。

四条肽链的N端，即L链的1/2与H链的1/4处（图7-2B）的氨基酸的排列顺序随抗体特异性不同而不同，故将该区称为可变区（variable region，

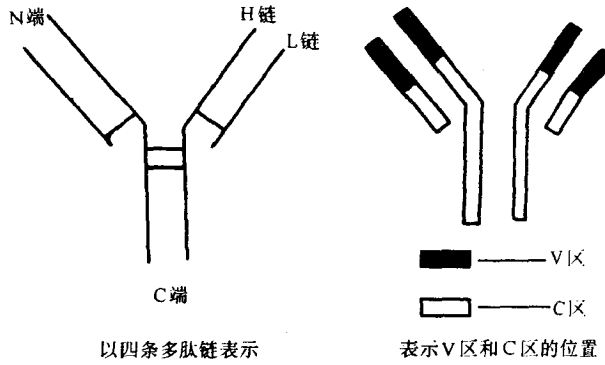


图 7－2　免疫球蛋白基本结构简易模式

A：以四条多肽链表示　B：表示 V 区和 C 区的位置

V 区），它赋予抗体以特异性，同时也是与特异性抗原结合的位点。一个 Ig 分子可与两个抗原决定簇结合。多肽链的 C 端即 L 链其余的 1/2 及 H 链其余的 3/4 部分的氨酸数量、种类、排列顺序及含糖量都比较稳定，故称为恒定区（constant region，C 区），Ig 可变区与恒定区的分界线在第 114 位氨基酸下，即 114 位以前的 N 端为可变区，115 位以后的 C 端为恒定区。两条相同的 H 链借二硫键相连，两条相同的 L 链也分别借二硫键连接在 H 链 N 端的两侧，因此，Ig 都是对称性的大分子。

由四条多肽链和几对二硫键连接成的一个基本单位为 Ig 单体。IgG、IgD、IgE 及大多数血清型 IgA 均为单体，而分泌型 IgA 为双体，即由两个单体 IgA 由 J 链相连，再加上一个分泌片段组成；IgM 为 5 聚体，即由 5 个单体 Ig 和 1 条 J 链相连而成（图 7－3）。

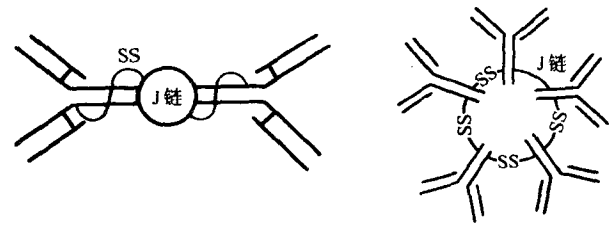


图 7－3　分泌型 IgA 和 IgM 分子模型

蛋白水解酶可将 Ig 分子裂解为不同大小的片段。如木瓜蛋白酶可将 Ig 分子裂解成两个抗原结合片段（antigen－binding fragment，Fab）和一个易结晶片段（crystallizable fragment. Fc），Fab 段的 H 链与 L 链由二硫键连接，若二硫键还原，

则形成 Fc 片段（图 7－4a）。胃蛋白酶可将 Ig 分子裂解成 F $(ab)_2'$ 和 Fc′两种片段。F $(ab)_2'$ 含有两个抗原结合部位。F $(ab)_2'$ 段 H 链片段的二硫键断裂，可形成两个 Fab′片段，Fab′段可进一步酶解成 Fv 片段。（见图 7－4b）。

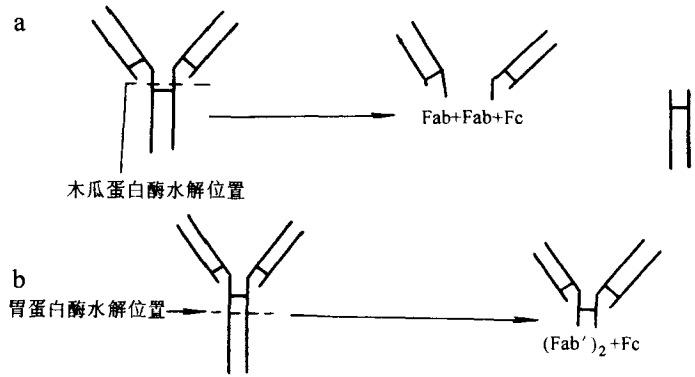


图 7－4 IgG 分子经胃蛋白酶和木瓜蛋白酶水解后的片段

在两条 H 链之间二硫键连接处附近的重链恒定区即在 CH_1 和 CH_2 之间有一个可以转动的绞链区（hinge region），当 Ig 与抗原结合时，该区可以转动，使可变区的抗原结合位点尽量与抗原配合和使不同距离的两个抗原决定簇结合，起弹性和调节作用，同时也使 Ig 变构，暴露出补体结合点。

Ig 的多肽链分子可折叠成几个由链内二硫键连接的球形结构，H 链有 VH、CH_1、CH_2、CH_3 四个球形结构区（IgM 和 IgE 还有 CH_4），而 L 链有 VL 和 CL 两个球形结构，VH 和 VL 是抗原结合部位；CH_1 是遗传标记所在，CH_2 是补体结合位点，CH_3 与细胞表面的 Fc 受体结合。

（二）免疫球蛋白的类别和血清型

理化特性及免疫学特性见表 7－2。

表 7－2 各类免疫球蛋白的理化特性及免疫学特性

性 状	IgG	IgA	IgM	IgD	IgE
分子量（KD）	150	170/390	900	170	200
重链类	γ	α	μ	δ	ε
亚类	γ1－2－3－4	α1 α2	μ1 μ2		
分子量（KD）	52	63	75	77	75

（续）

性　状	IgG	IgA	IgM	IgD	IgE
轻链型	κ、λ	κ、λ	κ、λ	κ、λ	κ、λ
分子量（KD）	22	22.5	22.5	22.5	22.5
基本四肽单位数	1	1～3	5	1	1
抗原结合价	2	2	5～10	2	2
占 Ig 总量（%）	75～80	7～15	5～10	0.3	0.00
生物学作用	抗菌、抗病毒、固定补体、通过胎盘与 SPA 结合	分泌型参与粘膜局部免疫、激活补体旁路途径	溶菌、溶血、固定补体	不明	介导Ⅰ型反应

（三）单克隆抗体

单克隆抗体（monoclonal antibody）是指由单个 B 细胞克隆生产的、在分子上是同质的抗体。

由于能产生抗体的 B 细胞在体外不能持续传代，而能在体外长期传代的骨髓瘤细胞则又不能分泌抗体，因此将能产生抗体的免疫淋巴细胞与骨髓瘤细胞融合，经过筛选、培养、克隆化，就可以建立既能分泌针对预防抗原的特异性单克隆抗体，又能在体外无限繁殖的杂交瘤细胞系，用来生产克隆抗体。该项技术称为淋巴细胞杂交瘤技术。

骨髓瘤细胞来自于人和动物的骨髓瘤细胞系，经过长期体外培养和药物筛选已失去了合成次黄嘌呤、磷酸核糖转移酶（HGPRT）或胸腺嘧啶核苷激酶（TK）的能力，本身也不能合成和分泌 Ig，将此骨髓瘤细胞与经预定抗原免疫的鼠淋巴细胞（通常用脾淋巴细胞）在聚乙二醇（PEG）等融合剂的作用下融合后，在含有次黄嘌呤（H）、氨甲喋呤（A）和胸腺嘧啶核苷（T）的选择性培养基（HAT）中培养，未融合的骨髓瘤细胞在培养基中 A 的作用下，DNA 合成的主要途径被切断，由于自身缺乏 HGRRT 酶或 TK 酶，也不能利用旁路途径合成 DNA，因此迅速死亡。不融合的淋巴细胞是短寿的，一般也只能存活几天。只有淋巴细胞与骨髓瘤细胞发生融合而形成的杂交瘤细胞能够存活下去，因为来自淋巴细胞的染色体弥补了骨髓瘤细胞失去合成 HGPPT 或 TK 酶的能力，而来自于骨髓瘤的染色体提供了能在体外长期连续传代的特性，部分杂交瘤细胞能分泌针对预定抗原的抗体，通过抗体检测可以把它们挑选出来，然后进行克隆化，即可建立能分泌针对预定抗原同质性抗体的杂交瘤细胞系（图 7－5）。

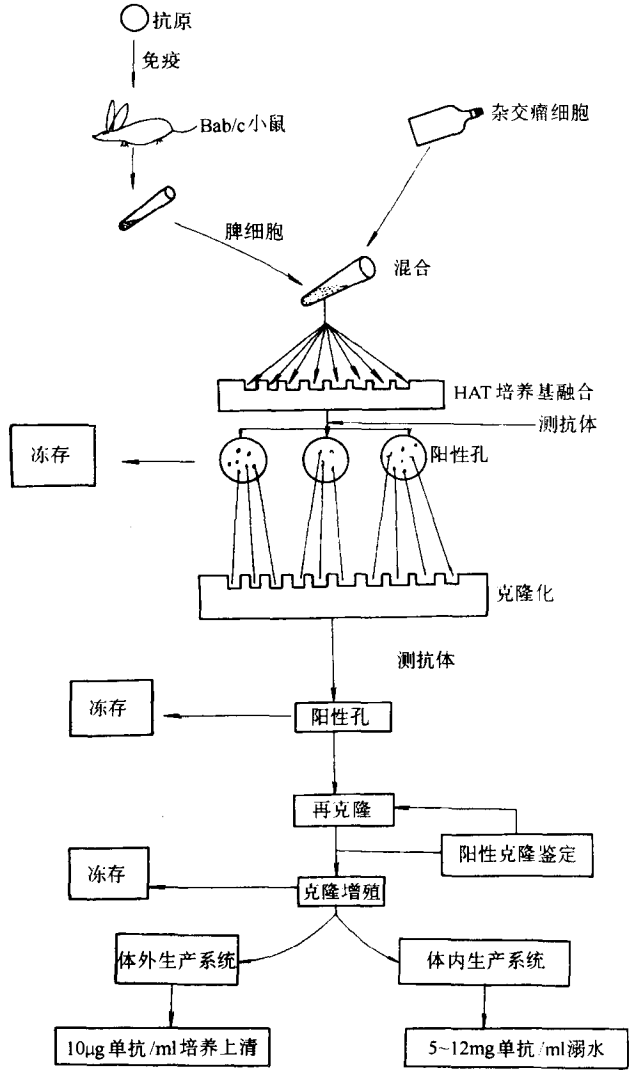


图 7-5 生产单克隆抗体的简易流程图

（四）基因工程抗体

基因工程抗体是将分子免疫学关于Ig基因结构与功能的知识与DNA重组技术、蛋白质工程相结合，按人们的意愿在基因水平上对Ig分子进行切割、拼接或修饰，人工合成、重新组装的新型抗体分子。它保留了天然抗体的特异性和主要生物学活性，去除或减少了无关结构，因此，比天然抗体更具有潜在的应用前景。

虽然单克隆抗体技术在基础理论研究和临床疾病诊断等多个领域的应用非常广泛，但由于鼠单克隆抗体的异源性而使其在临床治疗上受到限制。杂交瘤技术生产的单克隆抗体在制备和应用上也都有其难以克服的困难，例如，临床重复给药时体内产生人抗鼠抗体反应，使临床疗效减弱或消失。普遍认为，临床应用理想的单克隆抗体应是人源的，但人—人杂交瘤技术目前尚未突破，很难或不可能生产人杂交瘤单克隆抗体用于体内的免疫治疗。目前较好的办法是研制基因工程抗体，代替鼠源单克隆抗体应用于临床。

基因工程抗体兴起于 80 年代早期，其种类包括嵌合抗体、改型抗体、小分子抗体、Fab、单链抗体、单域抗体、抗体融合蛋白等（图 7－6）。

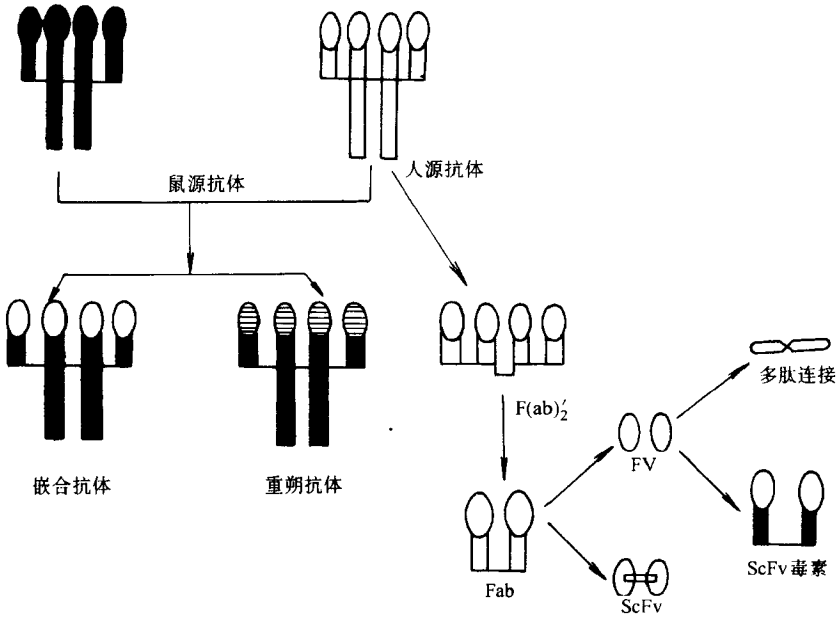


图 7－6　基因工程抗体示意图

构建基因工程抗体的基本步骤主要包括：（1）免疫球蛋白可变区基因的分离、克隆与序列分析；（2）抗体基因的修饰及表达载体的构建；（3）将修饰的抗体基因转染到能表达 Ig 基因的细胞中；（4）抗体基因的表达与筛选、鉴定。

二、抗体基因及其重组

免疫球蛋白基因由 3 组不连续的基因组控制，即 κ 轻链基因束、λ 轻链基因束和 H 链基因束。各个基因束中含有许多基因节段即基因子，在未分化的

胚系B淋巴细胞和其他有核细胞的染色体中，编码免疫球蛋白可变区（V区）和恒定区（C区）各个部分的基因节段是互相分离的。在B细胞的发育过程中，一部分基因节段互相接近，重新排列，形成功能性的免疫球蛋白基因单位，这一过程称为基因重排（gene rearrangement），重排的Ig基因经转录、转录后加工和翻译、翻译后修饰，成为功能性免疫球蛋白。人类免疫球蛋白H链基因、κ、λ基因分别位于第14、2、22号染色体上；小鼠免疫球蛋白H链基因、κ、λ基因则分别位于第12、6、16号染色体上。

（一）胚系B细胞中Ig轻链基因的结构

完整的轻链由3个基因片段编码：V和J基因片段编码可变区，C基因片段编码恒定区。

1.κ基因　人和小鼠Igκ链可变区前基因由多个基因节段组成。这些节段分为Vκ和Jκ链两组，κ链恒定区基因节段（Cκ）只含有单个外显子，位于Jκ下游，Jκ基因片段编码Ig轻链V区110个氨基酸的主要部分（1～95或96）。人和小鼠的Vκ基因节段至少有100种，并可根据核苷酸同源性将它们分为16个（小鼠）或4个（人）家族，各家族成员基本上在染色体上的一个区域内呈簇状串联排列（tandem array），其中许多是假基因（psuedogene）。但是，假基因可以通过突变、基因置换和不规则重组转录可能成为新的功能性基因，为可变区的多样性提供后备基因。每一个Vκ基因节段前面都有一个编码前导肽的前导序列，前导肽对轻链蛋的合成至关重要。

生殖系中Jκ基因片段位于Vκ基因片段下游，也呈串联排列，Vκ和Jκ中间有内含子分隔。Jκ基因编码轻链V区的其余部分（95或96～110），人和小鼠的Jκ基因各有5个基因节段，其中小鼠的$J\kappa_3$是假基因。

2.λ链基因　λ基因同样也分为3组：Jλ基因片段，Vλ基因片段和Cλ基因片段。Cλ基因片段中依次排列着各个亚型λ链的Cλ基因节段。

小鼠Vλ基因节段只有2个，可能产生的多样性远低于Vκ。由于λ链只占小鼠Ig轻链的5%，较少的基因节段并不影响轻链的多样性。小鼠Vλ、Jλ和Cλ基因节段的排列与λ基因节段的排列不同。每个Vλ与2个Jλ－Cλ相连，呈$V\lambda_1-J\lambda_3-C\lambda_3-J\lambda_1-C\lambda_1$和$V\lambda_2-J\lambda_2-C\lambda_2-J\lambda_4-C\lambda_4$，其中$J\lambda_4-C\lambda_4$是假基因（图7－7）。

人的Vλ前基因有30多个节段，它们也是串联排列，且每个Vλ基因节段的前面也都有一个前导序列。人Vλ至少有4～5个功能性基因节段和4个假基因节段，每个Cλ基因节段前面都有各自的Jλ基因节段。在人的Ig轻链中，λ链约占1/3，小鼠和人的λ基因在排列上的差异对人/鼠嵌合抗体的制备和应用不利。

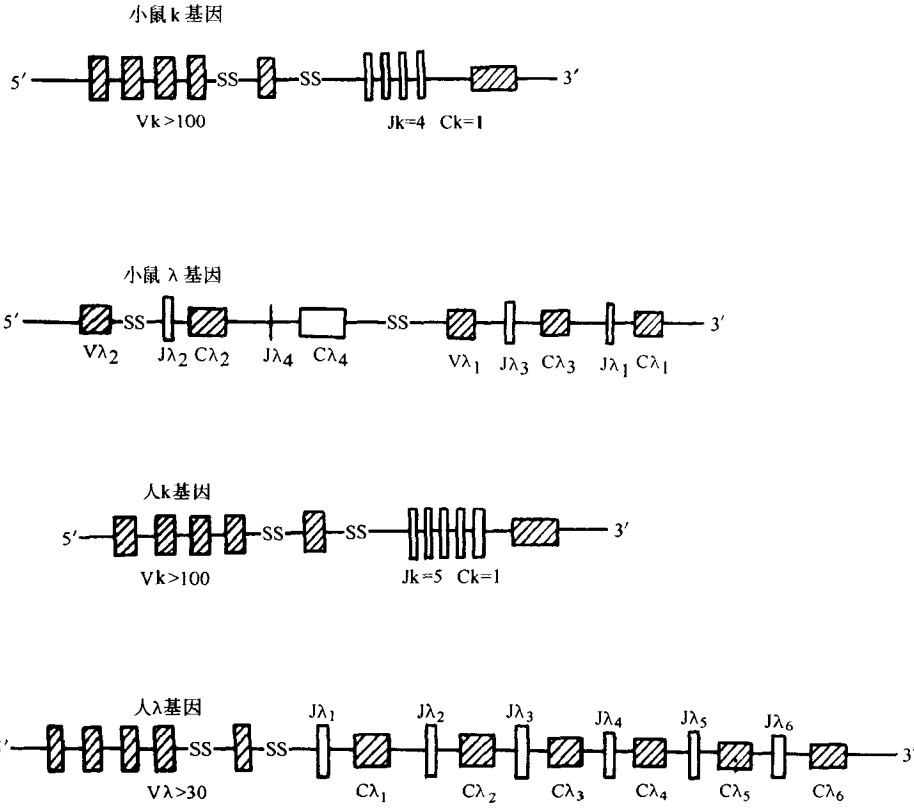


图 7－7　人和小鼠胚系 B 细胞中 Ig 轻链基因结构

（二）胚系 B 细胞中 Ig 重链基因的结构

重链基因结构主要有 3 个特点：（1）重链可变区由 V 基因节段、D 基因节段（diversity）和 J 基因节段 3 个基因节段区编码；（2）编码重链恒定区各亚类的 C 基因节段在可变区基因节段的下游成簇状串联排列；（3）每个亚类（δ 链除外）的重链恒定区基因节段上游都有一个转换区（switch region），基因节段中编码 Ig 各功能区的外显子之间有内含子分隔，在功能区外显子下游有控制 Ig 分泌或膜结合的元件。

1. 重链可变区基因节段

VH 基因节段　编码重链可变区的疏水前导肽和 N 端的 96～101 个氨基酸，其中包括互补决定区 1（complementary determining region 1，CDR_1）和互补决定区 2。每个 VH 基因节段有 2 个外显子，中间由约 100 个碱基对的内含子相隔，5′端外显子编码大部分前导肽，第二个外显子编码前导肽的后 4 个氨

基酸和可变区的其他部分。前导肽的作用是在合成重链时将重链固定在糙面内质网的膜上。

DNA序列分析和Southern印迹试验资料提示，小鼠的VH基因节段约有100～1 000个基因节段，分为13个家族，每个家族有4～100个以上的成员，各家族成员之间的核苷酸序列之间同源性很高，且呈簇状聚集，平均相距10～20kb。

Vκ基因节段区中也存在着假基因节段，按照基因节段中是否有可读框以及基因节段是否能重排成功能性基因这个标准，在人染色体上的95个VH基因节段中有44个没有ORF或无功能性重排的假基因节段。

DH和JH基因节段 在免疫球蛋白V区的CDR中，CDR_3多样性最显著，在抗原*l*抗体反应中也最重要。CDR_3由DH和JH基因节段区共同编码。DH节段族位于VH和JH基因节段族之间，长度约80kb（小鼠）。重排后的DH基因节段在序列和长度上变化很大，约编码1～15个氨基酸。目前证实的DH节段数目为15个（小鼠）或30个（人），也可分为3（小鼠）或7个（人）家族，各家族通常呈簇状分布。

JH节段族在DH基因节段区的下游，位于DH基因节段区上游约7kb处。人的JH有9个节段，其中3个是假基因，小鼠有4个节段。JH编码重链可变区的其余氨基酸部分。在B细胞成熟过程中，从VH、DH和JH基因节段库中各选择一个基因节段相互连接，重新排列成一个功能性的VDJ前基因，再与CH前基因连接。

2. 重链恒定区基因节段

重链恒定区基因节段 编码重链恒定区各亚类的基因片段（CH），在JH基因节段区的下游成簇状排列（图7－8），人的CH族中有3个假基因，$C\mu$和$C\delta$离JH最近，其次是各亚类$C\gamma$，最后是$C\alpha$和$C\delta$。整个重链恒定区分2～4个功能区和绞链区，在每个CH基因节段中同样含有分别编码各功能区的外显子。

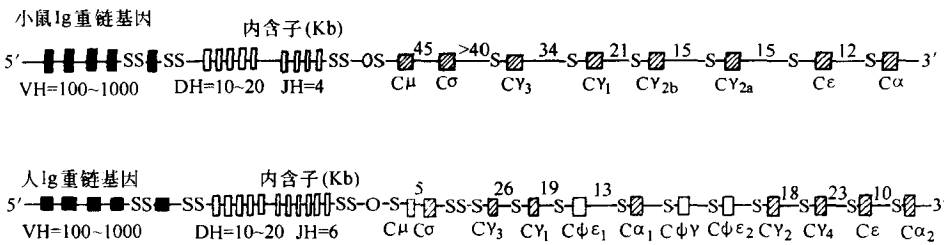


图7－8 人和小鼠胚系免疫球蛋白重链基因结构

■: V基因节段 ▯: J基因节段 ▨: C基因节段 ▥: D基因节段

□: 假基因 ○: 增强子 S: 转换位点

膜结合型 Ig 基因或分泌型 Ig 基因　理论上每一抗体都有膜结合型和分泌型两种形式。编码抗体膜结合区的基因片段和编码抗体分泌信号肽的基因片段位于重链恒定区基因片段的 3′端。

膜结合型 Ig 有一个穿膜的疏水区和一个伸入胞浆膜的尾肽序列。编码穿膜区和尾肽序列的基因位于最后一个功能区外显子的下游（图 7－8A、B），编码约 26 个疏水氨基酸的穿膜区和一个长短不一的亲水尾肽序列。μ 链的尾肽序列只有 3 个氨基酸，而 γ 链的尾肽序列有 26 个氨基酸。各种膜结合型 Ig 的穿膜区表现了高度同源性，在跨越膜双脂层时形成 α 家族的构型。这种同源性可使不同 Ig 与膜中同样的蛋白质相互作用而提供相同的活化信号。

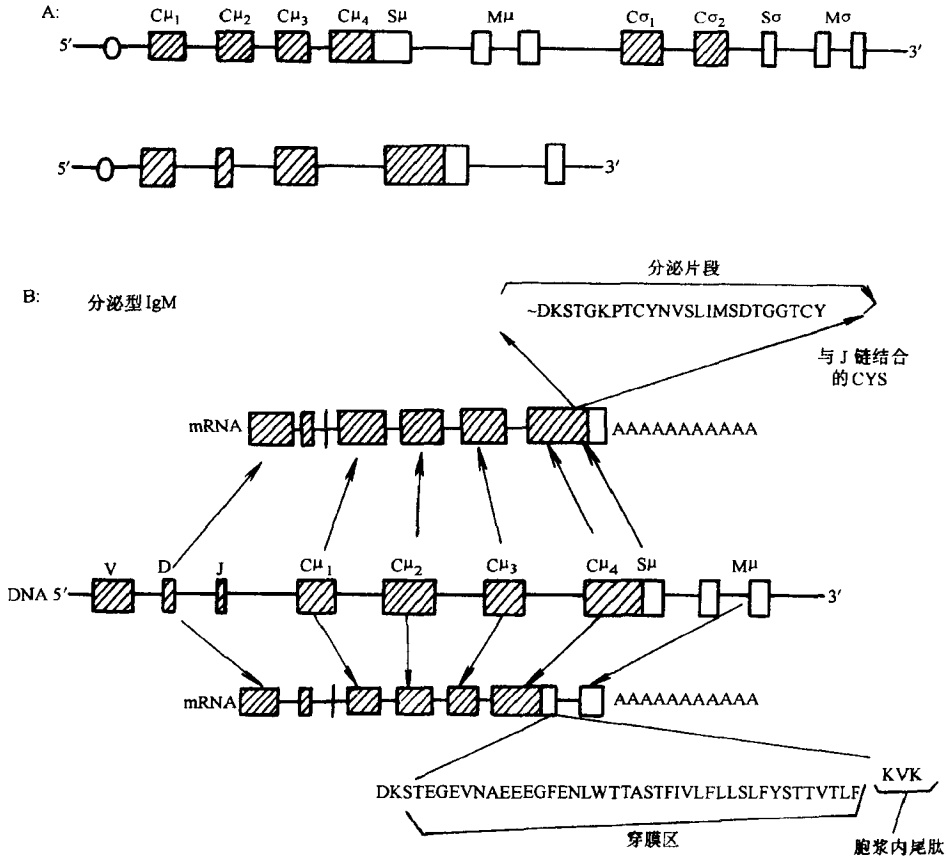


图 7－9　重链恒定区基因片段的结构以及分泌型的膜结合型 IgM 的合成

A. $C\mu$－$C\delta$ 和 $C\gamma$ 重链恒定区结构　1－4：编码各功能区的外显子　S：编码分泌信号肽的外显子　M：编码穿膜区和胞浆内尾肽的外显子　H：编码绞链区的外显子　B. 分泌型的膜结合型 IgM 的合成过程

分泌型 Ig 在合成时有一段插入内质网膜的信号肽。在除 Cδ 以外的其他重链恒定区基因片段中，编码信号肽的 DNA 序列紧接着 CH 的最后一个外显子；在 Cδ 基因片段中，编码信号肽的外显子位于 CH 下游，与 CH 的最后一个外显子不连续，有内含子分隔（图 7－9A、B）。

三、抗体基因重排与抗体多样性

（一）抗体基因重排的机制

1. 可变区基因重排的顺序　重链可变区（VH）基因片段和轻链可变区（VL）基因片段都是由多个节段组成。B 细胞在分化成熟的过程，从 VH 或 VL 的 V、D 和 J 基因节段库中各选择出一个基因节段，经 DNA 重排，与恒定区基因片段（首先是 Cμ）构成一个功能性基因进行转录，经转录后加工，再表达为免疫球蛋白（见图 7－10）。在 Ig 可变区前基因重排时，先分两个阶段构建重链可变区基因片段，即选择一个 DH 节段与一个 JH 节段连接在一起，然后选择一个 VH 节段与 DH－JH 相连。一旦 VH－DH－JH 重排成功并且形成的功能性重链基因得到表达，其产物就会抑制一条染色体上等位基因的重排。小鼠 VL 前基因重排时，κ 基因优先于 λ 基因。一个 Vκ 节段与一个 Jκ 节段连接成为功能性 Vκ－Jκ－Cκ 基因以后，其产物就会引发等位基因排斥，同时也会抑制 λ

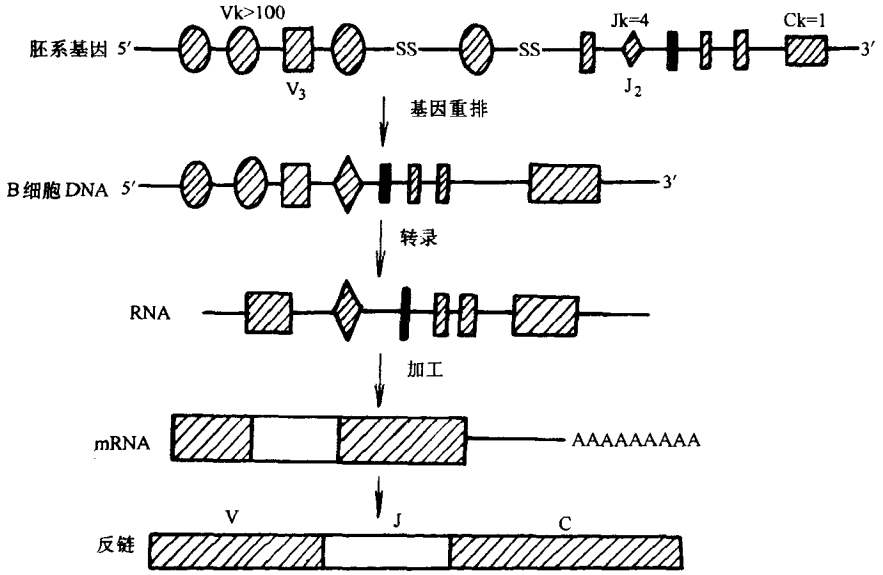


图 7－10　免疫球蛋白基因重排和表达示意图（以 κ 链为例）

基因重排，即所谓同型排斥（isotype exclusion）。在两条染色体上的κ基因重排都不成功时，才会启动λ基因重排。

如果一条染色体上重排的VH或VL基因片段不能获得功能性表达，即重排不成功，称为无产物重排（nonproductive rearrangement），细胞就会启动另一条染色体上等位基因重排。如果两条染色体上VH前基因都是无产物重排，或者κ基因和λ基因都是无产物重排，细胞即死亡。

2. *基因节段间连结的12－23bp规律*　VH、DH和JH以及VL和JL基因节段之间连接和精确机理还未被完全阐明。核苷酸序列分析发现，在每个VH和VL节段的下游和在每个JH或JL节段上游以及在每个DH节段的两侧，存在着一种特殊的、相互关联的侧翼序列或称共有序列（consensus sequence），这些侧翼序列能触发VDJ节段的重排并决定节段间的连接，所以又称为重组信号片段（recombination signal sequence，RSS）。每个侧翼序列含有一个7核苷酸的回文序列和一个9核苷酸的保守序列，两者之间有一段固定数目（12和23±1）的随机核苷酸序列分隔（图7－11B）。尽管在κ基因、λ基因和重链基因中，7核苷酸或9核苷酸序列都略有不同（图7－11A），但在被连接的两个基因节段之间，它们是反向互补的。这种结构意味着它们可以形成双股DNA，使两个待连接基因节段反绕成环状结构（图7－11C），在DNA双链修复系统的作用下，使两个基因节段互补连接在一起。细胞内DNA损伤修复系统在VDJ连接过程中的这种作用至关重要。已证明，在修复系统有缺陷的小鼠中，VDJ不能有效连接。

RSS的分隔7核苷酸和9核苷酸的随机核苷酸序列数目是固定的12bp或23bp。这个数目代表DNA双螺旋的一个或两个完整螺旋，说明围绕每个待连接基因节段的侧翼序列在空间上和方位上也是精确互补的，很可能因此决定或控制了重排的发生。研究发现，一个侧面有12bp分隔序列的基因节段只能与一个侧面有23bp分隔序列的基因节段重组或连接，反之亦然。这一现象称为12－23bp规律。因此，对重链来说，只要DH和JH以及VH和DH－JH之间能够重排，VH与JH之间就不会重排。7核苷酸序列和9核苷酸序列对重排基因节段的选择十分重要，其中7核苷酸序列的前三个核苷酸（CAC）是必需的。小鼠$J\kappa_3$节段CAC变成了TAC，所以不被重排，这是它成为假基因的重要原因之一。

解释Ig基因节段重排或连接机理的假说均以RSS和12－23bp规律为基础。

3. *VDJ与C的链接*　在B细胞的分化过程中，VDJ的重排及其与C的连接不依赖抗原刺激。此时，重排的VDJ与离它最近的CH基因（$C\mu$和$C\delta$）连接，并转录为长的RNA（VDJ－$C\mu$－$C\delta$）和短的RNA（VDJ－$C\mu$）。经转录后加工，长的RNA产生膜结合型和分泌型μ链mRNA或δ链mRNA；短的RNA产生膜结合型和分泌型μ链mRNA。此时细胞表达SmIgM和（或）SmIgD。因为$C\delta$基

因片段中没有转换区，所以 IgD 的表达与类型转换无关。

（二）形成抗体多样性的原因

免疫球蛋白多样性（immunoglubulin diversity）主要是指免疫球蛋白 V 区（VH 和 VL）的多样性，它使机体产生的免疫球蛋白能识别自然界中各种各样抗原性物质的不同表位。在基因重排时，功能性基因产生以后以及转录子在转录后加工过程中都可能产生多样性。多样性的产生至少与以下几个因素有关。

（1）组合性链接（combinatorial jointing） 免疫球蛋白的重链 V 区和轻链 V 区分别由 3 个（V、D、J）和 2 个（V、J）基因节段编码。在胚系中，VH、VL、DH、JH 和 JL 基因节段数目众多，使 B 细胞可以从中各选择一个基因节段组合并连接成各种不同的功能性基因。如小鼠，250～1 000 个 VH 节段、10～20 个 DH 节段、4 个 JH 节段、250 个 VL 节段、4 个 JL 节段可以组合成 10^7 以上的不同的功能性基因。这些功能性基因编码的重链可变区和轻链可变区，使免疫球蛋白可以识别 10^7 种特异性抗原。

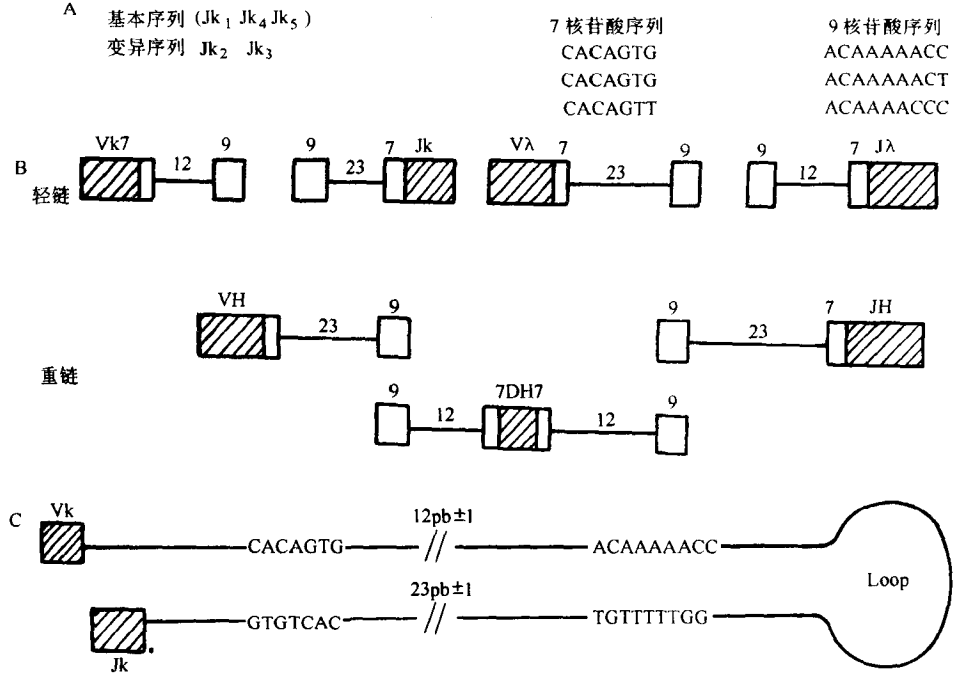


图 7－11 免疫球蛋白 VDJ 基因节段的重组信号片段及其连接

A.κ 链的 7 核苷酸和 9 核苷酸序列 B. 侧翼序列在 VDJ 基因节段的位置以及连接方式

C.RSS 连接时互补形成双股 DNA，在 DNA 修复系统作用下，V 和 J 基因节段相互连接

（2）连接多样性（junction diversity）　在DH－JH、VH－DH－JH和VL－JL连接时，由于连接过程的不精确性，大大增加了多样性，产生连接多样性方式主要有以下几种。

密码子错位：连接可以发生在连接部位的任何密码子之间，因此，同一对基因节段连接时，由于接合的核苷酸密码子不同，可以导致氨基酸序列中单个氨基酸残基的改变—增加或减少（图7－12A）。

移码突变：不精确的连接可引起移码突变。这种突变在连接完成后改变DNA的可读框（图7－12B），使连接点以后的氨基酸种类完全改变。研究发现，在连接点两边，各有约10个左右的核苷酸参与连接。如果在连接时有个别核苷酸缺失，就会使其后的全部氨基酸都发生改变，这种移码突变如果产生了终止密码或终止信号，就不能合成正常产物或只能合成无功能产物。

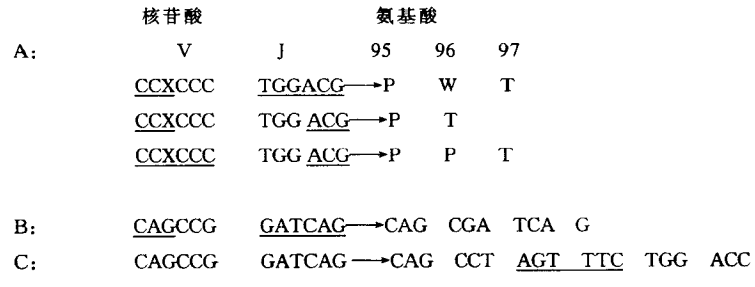


图7－12　连接多样性产生的几种形式

A.VJ连接的密码子变化使第三可变区95－97位氨基酸发生改变　B.移码突变，CG的缺失使整个可读框发生变化　C.N序列（AGTTTC）插入

N－序列插入：有时，在链接的过程中插入了一个并不属于待连接基因节段的外来核酸序列（N－序列，N－sequence）。N序列长度不等（约1～20个核苷酸），可以在V、D和J重排的任何位点插入，多见于CDR_3区。N序列的插入增加了可变区的长度，称为N区增加（N region addition，图7－12C）。多余的核苷酸能有效地改变免疫球蛋白结合抗原的特异性，但其发生率相对较低，N序列插入与V、D和J节段无关，而与末端脱氧核苷转换酶（TdT）的作用有关，在TdT缺陷小鼠中，可见Ig的CDR_3环式结构明显缩短。

（3）体细胞突变（somatic mutation）　也可以产生免疫球蛋白的多样性，插入突变和缺失突变可以发生在B细胞发育的任何阶段和Ig基因的任何部位。体细胞突变可以使鸡和兔的假基因节段转变为功能性节段，但未在其他动物中发现这种转变。

（三）抗体基因重排的控制

在胚系B细胞的免疫球蛋白基因重排时，发现了一些有规律的现象，例如，只有胚系B细胞才有Ig基因的重排，其他细胞不重排此基因；VDJ基因节段的重排按一定顺序进行，即DH与JH先连接，再进行VH和DH－JH的链接，在基因重排以后有等位基因排斥现象，即B细胞只表达一条染色体上的基因，产生一个有功能的重链（或轻链），另一条染色体保持胚系状态或是无产物重排。等位基因排斥保持了B细胞克隆识别抗原的单特异性。轻链基因重排后有同型排斥现象等，尽管目前对这些现象尚无满意的解释，但它们是确实存在的。

VDJ基因片段重排的启动需要重组活化基因（recombination activating genes，RAG）的表达。此基因的产物RAG_1和RAG_2介导VDJ重排起始步骤。增加内源性RAG_1和RAG_2基因转录时，VDJ重排的频率也伴随增加。RAG_1或RAG_2基因缺陷的纯合子小鼠，其T细胞和B细胞缺乏功能，也不发生VDJ重排。一旦VH基因节段重排成功并转录表达，RAG的表达就被抑制。其他与启动VDJ重排的有关基因产物还有LyF－1/1karo3、LEF－1、DCT－1和OCT－2，其中OCT－2是B细胞特有的。

在成功地重排了Ig基因以后，未成熟的B细胞在表面表达B细胞受体（BCR），即SmIgM和SmIgD，这种B细胞上的BCR如果针对自身抗原，在遇到自身抗原时即被排斥，而针对非己抗原的细胞，在遇到非己抗原时则被活化，此即克隆选择（clonal selection）。最近发现，BCR与自身抗原的相互作用还可以诱导第二次基因重排，即用新的VH节段代替原有的VH节段。其结果是改变了BCR的抗原特异性，同时也增加了BCR的多样性，这一现象称为B细胞受体编排（B－cell receptor editing）。BCR编排的机理仍需进一步研究。

免疫球蛋白的表达有组织特异性，它只在B细胞中表达，而且只在基因重排以后才表达。在每个VH和VL基因节段前面有一个转录启动子来启动重排基因的转录，V基因节段启动子位于转录启始位点上游90～120bp处，含有一个8核苷酸序列［TATAA（T）AA（T）］。RNA聚合酶Ⅱ识别这个TATA框，进而转录免疫球蛋白基因。启动子中另一个保守的15核苷酸序列［TGCAG（C）CTGTGNCCTG］在TATA框上游约60个核苷酸处。如果除掉此启动子，Ig基因就不能转录。V区启动子功能的发挥和Ig基因的组织特异表达还必须有增强子的参与，增强子序列位于重链基因或κ链基因的J和C之间的内含子中，小鼠α基因中未发现增强子序列，增强子序列作用于同一条染色体中的基因，其作用是顺式作用而不是反式作用。上述的组织特异性产物（RAG_1、RAG_2和LyF－1/1karos等）就是通过结合增强子和启动子来启动和控制Ig基因重排的。

已知κ链上至少有5个能与核因子（nuclear factor）结合的增强子序列，其中之一是含有8核苷酸的共有序列（OCTA），可以至少与2个核因子（OCT－1和OCT－2）结合。OCT－2是B淋巴细胞特有的，它决定了增强子元件的组织特异性，将这个OCTA人工插入珠蛋白基因中，珠蛋白基因即可在B淋巴样细胞中表达。因此，可能是B淋巴细胞特有的增强子蛋白作用于增强子序列和V区启动子，从而产生IgD组织特异性。

DNA重排对产生有转录活性的Ig基因至关重要。转录启始后，在重排的基因中仍然有间隔序列存在，必须在转录后除去这些序列才能进行翻译。小鼠κ链核RNA前体至少经两步加工才能成为mRNA，即将前导序列与V区连接，和将V区与C区连接，最后去除其中的内含子部分，mRNA经翻译后装配成为完整的免疫球蛋白。

B细胞转变为浆细胞以后，能高水平地表达免疫球蛋白。这种高水平的表达也受一些增强子的调节，已知在Cα的3′端有4个增强子，衔接排列在座位控制区，它们在浆细胞阶段活化，其中3α增强子（3E或hs1，2）是转录因子B细胞活化蛋白（DSAP）的结合位点，它的两端紧贴着8核苷酸序列和富含G的类似转换区的序列，分别结合不同的转录因子。在B细胞中，BSAP与hs1，2分离，使hs1，2和两个转录因子活化，进一步诱导类型转换和高水平表达免疫球蛋白。

（四）轻链和重链基因在转录后的加工与类型转化

VDJ与C的连接过程亦称为重链类型转换（hevey chain switch）。在受到抗原刺激以前，已重组的VDJ基因片段与Cμ片段和（或）Cδ基因片段一道已开始了转录，并可在细胞表面表达IgM和（或）IgD分子。受到抗原刺激后，业已重组的VDJ－Cμ或VDJ－Cδ片段与CH基因片段发生转换性重排，即VDJ从与Cμ或Cδ连接转换到与Cα、Cε或Cγ的连接，从而使B细胞产生具有同样抗原特异性（具有相同的VDJ）、但恒定区不同（Cα、Cε或Cγ不同）、能介导不同功能的其他类型的抗体（IgG、IgA、IgE）。

实验表明，类型转换由一系列称为转换区（switch region，S区）的重复序列介导，该序列始于CH基因片段的5′端，如与Cμ5′端连有一段重复序列［（GAGCT）n（GGGT）n］，每个重复中有20～80个核苷酸，重复数目约有2～5次，也有高达17次的。S区的侧面没有明显的侧翼序列。VDJ－CH的转换性连接可以在S区的许多位点发生，转换性连接后插入的串联重复序列在经过转录加工后被去除。

类型转换的机理可能是通过两种模式进行：（1）成环和删除模式，即含有C区（Cμ）的DNA序列在类型转换过程中形成环状结构，使重组的VDJ片段

能与新的C区靠拢。所有介于其间的DNA片段以及原始的$C\mu$基因片段均被删除，而某一特定的CH基因片段就能与VDJ片段相连接，经过转录而形成某一特定类型的免疫球蛋白。按CH基因片段在染色体上的排列顺序为$C\mu$、$C\delta$、$C\gamma$、$C\gamma_1$、$C\gamma_2$、$C\varepsilon$、$C\alpha$，依次形成不同的免疫球蛋白分子；（2）姐妹染色单体交换模式。在类型转换时，VDJ基因片段与新的CH基因片段之间的$C\mu$并没有被删除，而是在细胞分裂时通过重组方式转移到姐妹染色单体的对应的部位，$C\mu$可储存于胚系细胞内，且可被再次利用。

第三节 T细胞免疫反应

一、T细胞的类型及对外源抗原的识别

T细胞（thymus－dependent lymphocyte）是一种依赖胸腺的淋巴细胞，即其发育和分化依赖胸腺功能的完整。前胸腺细胞在胚胎发育期从卵黄囊或胚胎或成年期从骨髓进入胸腺，在胸腺微环境的影响下发育为成熟的T细胞。在此过程中，胸腺细胞须经过阳性选择和阴性选择：（1）阳性选择：前胸腺细胞最初为$CD_4^-CD_8^-$双阴性细胞，随后发育为$CD_4^+CD_8^+$双阳性细胞，后者表面的TCRα、β（T细胞受体）分子若能与胸腺皮质上皮细胞表达的MHC－Ⅰ或Ⅱ类分子结合，即可分别分化为CD_4^+或CD_8^+阳性细胞，否则就会发生细胞程序死亡，又称凋亡（apptosis）；（2）阴性选择过程：胸腺皮质和髓质交界处Mφ和树突细胞（DC）表达高水平的MHC－Ⅰ、Ⅱ类抗原，后者与自身抗原结合成复合物。经过阳性选择后的胸腺细胞能识别该复合物，即产生自身耐受并停止发育。不能识别核MHC分子—自身抗原复合物的胸腺细胞才能继续发育。经历上述与MHC有关的选择过程，T细胞才分化为能与异物抗原反应的CD_4^+或CD_8^+的单阳性细胞，随后经外周血循环进入外周淋巴器官中定居。

1.T细胞的表面标志 T、B细胞均属于小淋巴细胞，根据形态特征很难区分，T、B细胞表面有许多独特的膜分子，可作为其标记。

（1）T细胞受体（T cell receptor，TCR）是T细胞表面存在的能与抗原发生特异性结合的受体，故亦称抗原识别受体。大多数T细胞的TCR是由α、β两条链组成的异二聚体（heterodimer），每条链有2个功能区即V区和C区（图7－13），由α、β二聚体组成的TCR为TCR_2，分布在成熟的T细胞表面。人TCR_2分子量约为90KD，其α链为49KD，β链为40－43KD的糖蛋白，其细胞表型为CD_3^+、CD_4^+、CD_8^+。CD_3分子与α、β呈非共价键结合，形成完整

的 TCR－CD_3 复合体，共同参与对抗原刺激的免疫应答，二者缺一不可。TCR 识别抗原，由 CD_3 将抗原信息传递进入 T 细胞内使之活化。TCR_2 分为 CD_4^+、CD_8^+ 两个亚群。

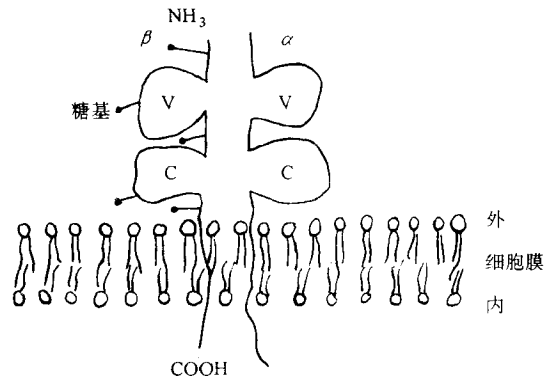


图 7－13　TCR 模式图

（2）T 细胞分化抗原（T cell differentiation antigen）是 T 淋巴细胞随着分化过程而出现或者发生变化的抗原，故称 T 细胞分化抗原。国际上统一以分化群（cluster of differentiation，CD）命名。这些膜分子与抗原的识别以及与 TCR 和其他 T 细胞膜分子互相作用过程有关，不同的 T 细胞或其亚群所表达的分化抗原种类或谱式格局也不相同。目前已鉴定的 CD 分子达 100 多种，CD 后面的数字表示不同的 CD 分子，如 CD_1、CD_2……等。T 细胞进入胸腺时不表达任何 CD 分子，开始分化时表达 CD_2、CD_7，进一步表达 CD_3，随之在同一细胞上表达 CD_4 和 CD_8，在胸腺髓质内 T 细胞分为两群，它们均有 CD_2、CD_3，但其中一群为 CD_4^+、CD_8^-，另一群为 CD_4^-、CD_8^+，这两个 T 细胞亚群可进入外周免疫器官（见图 7－14）。

（3）其他表面标记　其中之一是绵羊红细胞（SRBC）受体，又称 E－受体。人类 T 细胞与 SRBC 结合形成玫瑰花结，故此试验又称 E－玫瑰花结试验，（E－rosette test），用来鉴定 T 细胞或分离、提纯 T 细胞，该受体为 CD_2 分子。另一种是 Fc 受体（Fc receptor，FcR）。各类 Ig 的 Fc 受体已统一命名为 FcγR、FcεR、FcαR、Fcμ R 等，它们分别能与 IgG、IgE、IgA、IgM 等的 Fc 段结合。

2．T 细胞的亚群及功能　T 细胞按其表型及功能不同至少可以分为四个主要亚群（subset），即细胞毒性 T 细胞（cytotoxic T cell，CTL 或 Tc）、抑制性 T 细胞（suppressor T cell，Ts）、迟发性超敏反应 T 细胞（delayed type hypersensitivity T cell，T_{DTH}）、诱导－辅助性 T 细胞（inducer－helper T cell，T_i/T_H）及反抑制性 T 细胞（contrasuppressor T cell，Tcs）。这些 T 细胞的特点和

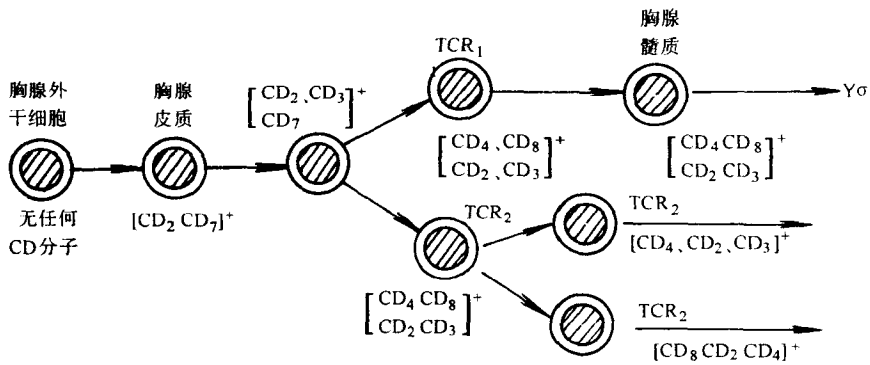


图 7－14 T细胞CD分子和TCR的表达

功能总结于表7－3。CD_2 和 CD_3 是所有T细胞共同的表面标志，CD_4 和 CD_8 分别存在于 T_i/T_H 和 T_s/T_c 表面，只有表达 CD_4 或 CD_8 分子才表明T细胞进入成熟阶段。严格地说，鉴定成熟T细胞还应检测TCRα、β分子。

表 7－3 人T细胞亚群表型及功能

T细胞亚群	表 型	MHC识别限制性	功 能
T_c 细胞	CD_3^+、CD_4^-、CD_8^+、CD_{11}^+	Ⅰ类分子	杀伤靶细胞分泌IFN－Y
T_s 细胞	CD_3^+、CD_4^-、CD_8^+、CD_{11}^+		抑制靶细胞功能分泌TSF
T_H 细胞	CD_3^+、CD_4^-、CD_8^+、CD_{cu8}^-细胞	Ⅱ类分子	
T_I 细胞	CD_3^+、CD_4^-、CD_8^+、CD_{cu8}^+		诱导 T_H 和 T_s 成熟
T_{DTH}细胞	CD_3^+、CD_4^-、CD_8^+	Ⅱ类分子	参与迟超反应，分泌MIF、IFN－Y等LF
反抑制性T细胞	CD_3^+、CD_4^-、CD_8^+		分泌 $T_{cs}F$ 对抗 T_s 功能
(T_{cs})			使 T_H 恢复活性

二、T细胞受体基因

T细胞受体是T细胞识别抗原肽－MHC分子复合物的表面成分，属于免疫球蛋白超家族，组成TCR亚单位的有4个，即α、β、γ、δ。它们分别由不同的基因编码。与Ig基因一样，TCR的可变区和恒定区基因片段也必须在胚系T细胞中经历基因重组才能形成功能性基因。编码V功能区者为V基因片段，其中又可再区分为若干个节段，编码C区者为C基因片段，其他还有J和D基因节段。这四类基因片段分别位于不同的染色体上，人的α、δ基因位于第14对染色体上，β、γ基因位于第7对染色体。

1.α基因　由Vα、Jα和Cα3个基因节段组成，各节段均含有多个基因片段。人和小鼠的α基因约含有50～100个Vα片段，这些节段至少可以分10～20个家族，每个家族有1～10个成员，族间的同源性为35%，Jα位于Cα基因片段的5′端，全长约100kb，比其他TCR基因或Ig基因要大得多。Cα只有一个节段，包括4个外显子，分别与多肽链不同功能区氨基酸的转录有关。其中第一外显子编码C功能区的胞外部分，第二外显子编码连接多肽链与细胞膜的连接肽；第三外显子则编码跨膜部分及胞质内短尾；第四外显子编码3′末端非翻译部分。在Cα3′端还有T细胞特异性转录增强子和沉默子调控元件。

2.β基因　TCRβ链基因与Ig重链基因相似，由Vβ、Dβ、Jβ和Cβ4个基因节段组成。小鼠的Vβ区节段由25个片段组成17个亚簇，人的Vβ区约有50～100个节段，分为30个亚簇，各Vβ节段随机排列，位于D、J基因节段上游，但Vβ14（鼠）、Vβ30（人）靠近3′端位，位于D、J、C下游。

小鼠Dβ有2个成员，即$D\beta_1$和$D\beta_2$。$D\beta_1$位于Vβ节段下游，在Jβ节段族5′端的500～600pb外，Jβ基因节段簇则位于Dβ和Cβ之间，共有7个片段。每一个Jβ片段编码V区中的15～17个氨基酸。小鼠和人类的Jβ节段中有6个属于功能基因节段，与基因的转录有关。Jβ的最后一个片段$J\beta_7$与Cβ的第一个外显子$C\beta_1$紧密相连。

小鼠$C\beta_1$和$C\beta_2$基因节段高度同源，编码产物中仅有4个氨基酸不同，在人类中亦是如此，仅有6个氨基酸不同。两种基因片段均有相似的外显子、内含子结构。每种Cβ基因节段都含有4个外显子，第一外显子编码细胞外功能区以及连接细胞外功能区与跨膜部分连接肽中的一部分，连接肽的其余部分则由第二外显子和第三外显子的一部分编码。第三外显子的其余部分则编码跨膜部分。第四外显子编码胞质内短肽（有5～6个氨基酸）以及mRNA3′端非翻译部分。目前未发现Cβ基因中有与分泌相关的片段，说明天然的TCRβ链都是膜结合的。TCRα、β基因的主要特征列于表7－4。

表7－4　TCRα、β基因的主要特征

基　因	人　类		小　鼠	
	α基因	β基因	α基因	β基因
基因节段长度	～1 000bp	～900bp	～900bp	700～800bp
V基因节段长度	100bp	75～150bp	100bp	25bp
D基因节段长度	0	2bp	0	2bp
J基因节段长度	50～100bp	13bp	49～50bp	14bp
C基因节段长度	1kb	2kb	1kb	2kb
TCR谱		$\sim 10^{15}$		$\sim 10^{10}$
重　排	V－J	V－D－J	V－J	V－D－J

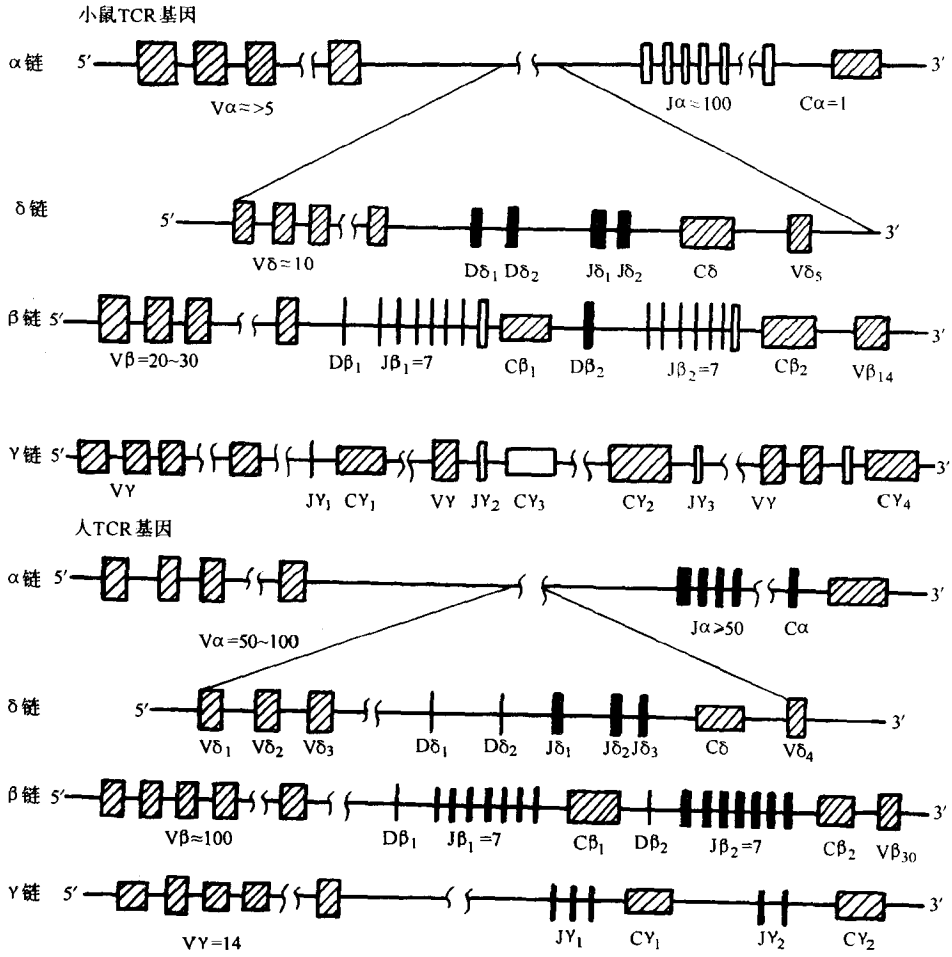


图 7－15　小鼠和人 TCR 基因结构

3.γ 链基因　γ 链由 Vγ、Jγ 和 Cγ 三组基因节段编码，未发现有 D 基因节段存在。目前已鉴定的人类 Vγ 基因节段有 14 个，其中 8 个是功能性基因节段，6 个是无功能的假基因节段，共分 4 个亚族，最上游的 9 个 Vγ 具有高度同源性（约 76%～91%），归于 VγⅠ组，其中有 4 个假基因节段；VγⅡ－VγⅣ位于 VγⅠ的下游，各有一个功能基因节段和两个假基因节段。Jγ 基因节段有 5 个，分为 $J\gamma_1$ 和 $J\gamma_2$ 两组，$J\gamma_1$ 含有 3 个节段；$J\gamma_2$ 含有 2 个节段，位于 $C\gamma_2$ 的上游（但 $C\gamma_1$ 位于两组之间）。由于只有 4 个 Vγ 和 5 个 Jγ 功能性胚系基因节段，故 TCRγ 的多样性是有限的。

在每一组 Jγ 链基因节段的下游均有一个 C 区基因片段。人类 $C\gamma_1$ 基因片

段所编码的连接肽均含有半胱氨酸残基，这种氨基酸残基与δ链或另一γ链形成链间二硫键至关重要。人类Cγ基因节段中有3个外显子，第一个外显子编码细胞外功能区；第二外显子编码连接肽；第三外显子编码连接肽的其余部分、跨膜部分、胞质内尾肽以及mRNA3′端非翻译部分。

4.δ链基因　δ基因的定位有其独特之处，即此类基因位于Vα和Jα基因节段区之间，Vδ基因节段位于Vα节段的3′端，V区中包括Dδ和Jδ。目前已鉴定的δ基因含有6个Vδ、3个Dδ，3个Jδ和1个Cδ；Vδ中的5个（$V\delta_1$—$V\delta_5$）位于Cδ上游，另一个（$V\delta_6$）则位于Cδ的下游，Vδ与Vα基因节段区间有氨基本酸序列同源性，如$V\delta_4$与$V\alpha_6$有96%的同源性。$V\delta_1$既能在δ位点重排，也可能在α位点重排，Vδ和Vα基因节段间可能有部分重叠。

δ基因位于α位点的这一情况可能具有潜在意义，因为如此，Vδ基因节段既可在δ位点重排，也可以在α位点重排，从而增加δ基因多样性。

三、TCR基因的重排及其多样性产生的机制

TCR基因重排在胸腺中完成。T细胞中TCR基因的重排顺序与B细胞中Ig重排顺序相似。在前T细胞中，TCR的β链VDJ基因节段的D和J首先连接在一起，然后在胸腺细胞中再与V连接，形成VDJ基因节段。VDJ基因片段或VJ基因片段（α链和γ链）重排完成后，再与C基因节段连接成为功能性基因。TCR基因重排的顺序在胸腺中受到严格的控制。

TCR可变区重排的机理与Ig可变区基因片段重排机理相似。在四种TCR基因（α、β、γ、δ）的VDJ基因节段的侧翼都有重排信号序列（RSS），它们之间的连接也遵循12－23bp规律，通过7、9聚体的互补而形成茎状和环状结构，使V－J序列靠拢，然后可能在重组酶的作用下而连接，去除茎环序列，形成重排的新DNA序列。

完整的TCRα、β分子的合成首先必须要有Vα和β基因节段的重排，形成功能性V基因片段。Vα基因片段由Vα、Jα基因节段重排后形成，Vβ则先发生Dβ－Jβ重排，然后发生Vβ－Dβ－Jβ重排。在V－J、D－J或V－D－J重排过程中，各基因节段的相互重排连接过程中并非完整无误，有时在重排过程中，可以从编码基因的一端或两端缺失几个碱基，或插入非胚系编码核苷酸（non－germ line encoded nucleotide），进一步增加重排的多样化，使V（D）J基因产生功能性和非功能性的两种重排。资料表明，发育中的T细胞可以通过等位基因排斥机制确保T细胞只能表达出来自一个等位基因重排编码的克隆特异性TCRα、β分子。

TCR多样性产生的机制包括：（1）基因组中有较多编码V功能区的基因片段；（2）V区由V、D和J片段组成，它们能互相结合而形成功能性基因；（3）β基因和δ基因有多种重组方式，如V－D－J、V－J、V－D－D－J；（4）重排中V区基因片段组合方式的多样化更可促使TCR多样性程度的增加，如V－D、D－J或V－J两片段的结合可在基因组中的任何一处发生。连结过程可发生移码，这种现象多见于TCRβ基因和δ基因的D片段。额外核苷酸可插入连结部位，如非胚系碱基的插入能增加重排的多样性；（5）α链和β链或γ链也可以与另一γ链连结（γ—γ），形成TCR肽链的抗原结合部位。虽然突变因素不能被看作增加TCR多样性的机制之一，因为在胚系细胞表达的V区基因节段中并未显示出有点突变，但是TCR基因的突变也不能完全否定。不过突变的发生对于T细胞的抗原识别过程是不利的。

第四节　主要组织相容性复合体

20世纪初发现，在不同种属或同种不同系别的个体间进行正常组织移植时会出现排斥反应，这是供者与受者组织不相容的表现。后来证实这实质上是一种免疫反应。机体内与排斥抗原有关的抗原系统多达20个以上，主要位于细胞表面，称为移植抗原或组织相容性抗原（histocompatibility antigen）。凡能引起强而迅速排斥反应的抗原称为主要组织相容性抗原，引起较弱排斥反应的抗原称为次要组织相容性抗原，它们都由体细胞的基因编码。主要组织相容性抗原包括多个系列，组成复杂的抗原系统，编码这一系统的基因位于同一染色体片段，是一组紧密联锁的基因群，称为主要组织相容性复合体（major histocompatibility complex，MHC）。在不同种属的哺乳类动物中，MHC编码的抗原系统有不同的名称，但它们的组成、结构、分布和功能等却很相似。控制免疫应答能力和免疫调节功能的基因即免疫应答基因（immune response gene，Ir gene）也位于MHC中，因此，MHC不仅与移植排斥反应相关，而且也与免疫应答和调节有关。小鼠MHC称为H_2复合体，人的MHC称为HLA复合体。

一、小鼠主要组织相容性复合体

1．小鼠MHC－H_2的结构和分类　小鼠MHC－H_2复合体位于第17号染色体上，可分为3类4区，其中Ⅰ类基因包括K区、D区、L区，Ⅱ类基因为i区，Ⅲ类为S区。

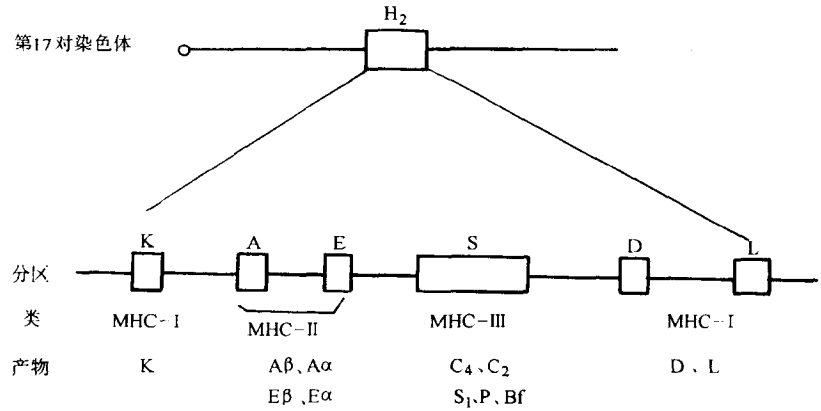


图 7－16　小鼠 H_2 复合体结构

Ⅰ类基因中，K、D、L 分别编码单一肽链 K、D、L，它们和 β2－假球蛋白结合后在细胞表面表达Ⅰ类分子，i 区中的 A、E 主要表达 Aα、Aβ、Eα、Eβ，分别为 H_2 复合体Ⅱ类基因ⅠA（表达 Aα/Aβ 异二聚体）和ⅠE（表达 Eα/Eβ 异二聚体）编码，i 区基因主要通过控制 TH 细胞的激活而影响抗体的生成，从而调控机体的免疫应答能力，所以被称为免疫应答基因（Iγ 基因），所表达的产物为免疫相关抗原（immune－associated antigen），简称 Ia 抗原。

S 区主要编码补体成分 C_2、C_4、Bf 等。

2. 小鼠 H_2 的Ⅰ类分子和Ⅱ类分子特点　表 7－5 列举了小鼠 H_2 Ⅰ、Ⅱ类基因产物的结构、分布和功能特点，从中可以看出Ⅰ类分子由 K 区或 D、L 区编码的重链和非 H_2 基因编码的 β_2（12KD）所组成，主要参与移植物急性排斥、抗体产生、细胞裂解和向 CD_8^+ T 细胞呈递抗原；而Ⅱ类分子两条链分子量相近，α 链、β 链分别为 35KD 和 28KD，主要分布在 B 细胞、巨噬细胞和激活 T 细胞表面，参与移植物慢性排斥、抗体产生、移植物抗寄主反应（GVHR），激发混合淋巴细胞反应（MLR）和向 CD_4^+ T 细胞呈递抗原。

表 7－5　小鼠 H_2 Ⅰ类和Ⅱ类分子的特点

特　性	Ⅰ类（K、D 区产物）	Ⅱ类（Ⅰ区产物）
化学组成	糖蛋白（45KD）+ β_2 激球蛋白	35KD 糖蛋白 + 28KD 糖蛋白
组织分布	普遍存在	B 细胞、巨噬细胞为主

（续）

特　　性	Ⅰ类（K、D区产物）	Ⅱ类（Ⅰ区产物）
参与移植排斥	急性作用强	慢性排斥
产生抗体	++++	++++
刺激 MLR	+	++++
裂解细胞	++++	+
抗原呈递	激活 CD_8^+ T 细胞	激活 CD_4^+ T 细胞

二、人类 MHC－HLA 复合体

人类主要组织相容性复合体（MHC）称为人类白细胞抗原（human leucocyte antigen，HLA），是因为该抗原首先在白细胞表面发现的。该基因区位于第 6 号染色体短臂 6p21.3 区域，到目前为止，已被命名的 HLA 基因座位数已达 129 个，能表达产物并显示功能的约占一半，主要分成 3 组：HLAⅠ、HLAⅡ、HLAⅢ。

1.HLAⅠ、Ⅱ类基因　HLAⅠ类基因从第 6 号染色体中心一侧开始依次为 HLA－3、－C、－E、－A、－G、－F 及Ⅰ类样基因或基因片段 HLA－17、－X、MHCⅠ类链相关基因 C、－30、－L（92）、－J（59）、－80、－21、－K（70）、－16、－H（54）、－90、－75、MCCD 等共 21 个基因。目前，除 HLA－A、B、C 基因和一部分 HLA－E、F、G 的功能外，其他Ⅰ类基因的功能知之不多（图 7－17）。

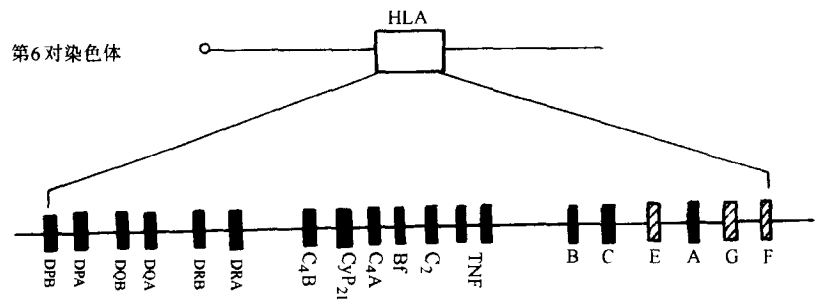


图 7－17　人类 MHC 结构

Ⅰ类抗原由非共价连接的两条多肽链 α 链、β 链组成，α 链由第 6 号染色体上Ⅰ类基因编码，β 链由第 15 号染色体上基因编码。HLA－B，C、－A 抗原除红细胞和精子外几乎存在于所有的有核细胞，而 HLA－B、－E、－G、－F 在组织发生时期的细胞中进行多特异性地表达。编码 α 链的Ⅰ类基因约有 7 个内含子和 8 个外显子（见图 7－18），其大小约 3.5kb，第Ⅰ外显子编码前导序列，第 3、4 外显子编码 α 链的 α_1、α_2、α_3 区，第 5 外显子编码连接多肽和膜贯通蛋白，余下的第 6、7、8 外显子主要编码细胞内区域和非翻译区域蛋白。HLAⅠ类抗原的多态性主要由 α_1、α_2 区的第 2、3 外显子决定。不同的Ⅰ类抗原分子与内源性抗原多肽，如病毒、癌抗原、寄生虫等（多为 8～9 个氨基酸）结合，形成 HLA/多肽复合体，提呈到 CD_8^+ T 细胞受体。HLA－Ⅰ类基因的多态性非常多，到现在为止，已发现 HLA A 等位基因 61 个，HLAB 有 136 个，HLAC 有 37 个，HLA E 和 HLA G 各 4 个。预计今后还会发现新的等位基因。

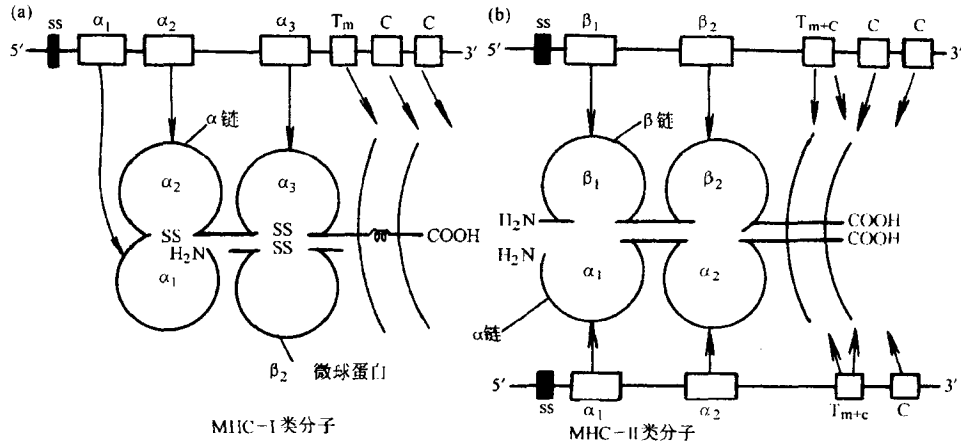


图 7－18　HLA－Ⅰ类和（a）Ⅱ类（b）分子及其编码的基因

2.Ⅱ类基因　HLAⅡ类基因从中心一侧开始依次为 DPB、DPA、DQB、DQA、DRB、DRA 基因，其中 A 代表 α 基因，B 代表 β 基因。Ⅱ类抗原是由第 6 对染色体上Ⅱ类基因编码的 α 链与 β 链以非共价键的形式结合的异源二聚体分子。HLA DR、DQ、DP 抗原在巨噬细胞、B 细胞、朗格罕细胞、树突状细胞及活化的 T 细胞等免疫系统的细胞中表达。编码 α 链的Ⅱ类基因有 5 个外显子，其大小约 6kb。第 1 外显子编码主导序列和 α_1 区的一部分，第 2、3 外显子编码 α 链的 α_1、α_2 区。第 4 外显子编码连接多肽和膜贯通蛋白的一部分。第 5 外显子主要编码细胞内区域和非翻译区域。编码 β 链的Ⅱ类基因有 6 个外

显子，其大小约为 8kb，其编码的顺序与 α 链相同，HLA Ⅱ类抗原的多型性主要由编码 α、β_1 区的第 2 外显子，尤其是 β_1 区的第 2 外显子决定的。不同的 Ⅱ 类抗原分子与外源性抗原多肽（多为 13～25 个氨基酸）如细菌多肽结合，形成 HLA/多肽复合体，提呈到 CD_4^+ T 细胞受体。详见表 7－6。

表 7－6　HLA Ⅰ类、Ⅱ抗原分子的结构、组织发布及功能特点

命　名	分子结构	抗原肽结合结构域	组织分布和表达	功能及其对特异性免疫应答的约束作用
Ⅰ类　HLA－A B、C	α 链（45KD） β 链（12KD）	$\alpha_1+\beta_2$	所有有核细胞表面	识别和呈递内源性抗原肽与共受体 CD_8 分子结合，约束 CTL 效应功能
Ⅱ类　HLA－DR DQ、DP	α 链（35KD） β 链（28KD）	$\alpha_1+\beta_2$	B 细胞、APC 激活的 T 细胞	识别和呈递外源性抗原肽，与共受体 CD_4 分子结合约束 Th 细胞识别功能

3. Ⅲ类基因　位于Ⅱ类与Ⅰ类基因区域之间，主要有补体 C_2、C_{4A}、C_{4B} 基因，B 因子（Bf）基因，21 羧化酶基因（CyP 21B）、CyP 21A 基因、热激蛋白（HSP70）基因、肿瘤坏死因子基因（TNF－α）TNF A、TNF B 和淋巴毒素基因（LTA）以及功能不明的 XA、RD、G11 基因等，这些基因都属于与 HLA 无关的非 HLA 基因，习惯上列为 HLA－Ⅲ类基因。

三、MHC 的生物学意义

70 年代以前已发现异体器官或组织移植时，MHC 作为主要的移植抗原可以引起强烈的宿主抗移植物反应或移植物抗宿主反应，自 1974 年 Doherty 和 Zinkergenal 发现自身 MHC 分子与 T 细胞识别外源性抗原有关后，才逐步认识到自身 MHC 分子在免疫应答过程中和免疫系统的发育中的重要作用。

1. MHC 在移植反应中的作用　同种异体组织移植时，若供、受体移植抗原不同，将会诱发受体产生明显的移植排斥反应。实验表明，供、受体的Ⅱ类分子不同时，供体Ⅱ类抗原能直接刺激受体 CD_4^+ T 细胞的活化及淋巴因子的分泌成熟，导致移植物的破坏。所以，MHC Ⅱ类抗原错配启动了免疫应答，而Ⅰ类抗原错配是导致免疫效应阶段被攻击的靶子。因此，临床上器官和骨髓移植时，要首先选择与受体 HLA Ⅰ类和Ⅱ类匹配的供体，其次选择Ⅱ类抗原

相配的供体，Ⅱ类抗原的匹配比Ⅰ类抗原的匹配更为重要。

2.MHC 在 T 细胞的分化成熟中的作用　MHC 分子参与胸腺细胞分化和成熟的过程。在阳性选择中，MHC－Ⅰ类分子使 CD_4^+、CD_8^+ 双阳性 T 细胞分化为 CD_8^+ 单阳性 T 细胞，而 MHC－Ⅱ类分子使双阳性 T 细胞分化为 CD_4^+ 单阳性 T 细胞，即 T 细胞两大亚群的分化过程有赖于它们对两类 MHC 分子的识别。在阴性选择中，Ⅰ类分子与自身多肽的复合体可诱导对自身应答的 CD_8^+ T 细胞凋亡，而Ⅱ类分子与自身多肽的复合体可使相应的对自身应答的 CD_4^+ T 细胞凋亡，这有利于非己抗原应答的 T 细胞发育成熟。在自身 MHC 分子诱导下，在胸腺发育的 T 细胞获得了识别自身的与非己抗原的能力，并对它们发生不同的应答。

3.MHC 对 APC 抗原呈递的作用　T 细胞通常只识别抗原呈递细胞（APC）呈递的由抗原肽和 MHC 分子组成的复合结构。这一识别是通过 T 细胞和 APC 之间的“TCR－肽－MHC”三元体或三分子复合物来完成的。这里，TCR 是 T 细胞接受抗原刺激的受体，其配体是抗原肽加上 MHC 分子，构成所谓的双重识别，即同一受体（TCR）的 α 链和 β 链分别以各自的第 1 和第 2 互补决定区识别 MHC 分子，以及第 3 互补决定区识别位于 MHC 分子抗原结合沟槽中的抗原肽，互补决定区（complementarity－determining region，CDR）通常位于高变段，代表了 TCR 和配体（此处是抗原肽和 MHC 分子）相结合的关键部位。与此同时，T 细胞表面的共同受体（coreceptor）CD_4 分子（单链）和 MHC－Ⅱ类分子非多态性的 α_2/β_2 部分结合，CD_8 分子（双链）和 MHC－Ⅰ类分子 α_3 结构域结合，结果，Ⅱ类分子参与 CD_4^+ Th 细胞识别抗原，而Ⅰ类分子参与 CD_8^+ Th 细胞识别抗原。

抗原从完整的分子转变成能被 TCR 识别的小肽并和 MHC 分子相结合，这个过程在 APC 中完成，称为抗原的加工和呈递，是免疫识别过程中的重要事件，也是 MHC 发挥功能和参与免疫调节的关键步骤。MHC 的功能是在细胞内捕捉相应抗原肽并将它转送到细胞表面、以肽－MHC 分子复合物形式提供给 TCR 识别。抗原的加工和呈递有两条主要途径，即外源性抗原加工途径和内源性抗原加工途径。这里的外源与内源是指抗原来自细胞外还是细胞内，而不是指体内或体外。

外源性蛋白质抗原（如菌体蛋白）经 APC 胞饮后进入溶酶体，蛋白质抗原被溶酶体酶解成小的抗原片段（抗原肽），同时内质网腔中合成 MHC－Ⅱ类分子（异二聚体）以及 γ 链，形成 α/β/γ 三链聚合物，移向溶酶体。α/β 进入溶酶体后 γ 则被解离，再与抗原肽结合，所形成的抗原肽－MHCⅡ被送到细

胞表面，由 CD_4^+ Th 细胞的 TCR 识别。由此启动免疫应答。

内源性蛋白质抗原（如病毒）直接进入胞质溶液，被低分子量多肽（LMP）捕获、酶解成小肽，在 HSP70 引导下通过内质网上的特定结构进入内质网腔，再经由腔内一个称之为 9p96 的蛋白质引导和修饰，抗原即成为 8～10 肽大小，进入腔内新合成的 MHC－Ⅰ类分子的肽结合沟槽内。Ⅰ类分子构型发生变化，迅速通过高尔基复合体被转送至细胞表面，被表达 CD_8 分子的 Tc 细胞所识别。

4.MHC 的限制性　T 细胞在识别细胞表面抗原决定簇的同时，还必须识别细胞上的 MHC 分子，具有同一 MHC 表型的免疫细胞间才能有效地相互作用。这一现象称为 MHC 限制性（MHC restriction）。

T 细胞的 TCR 是由 α/β（或 γ/δ 链）组成的异二聚体构成，发育过程中发生 V、CDJ、C 区不同基因片段的重排和拼接，最终合成结构连续的受体分子。由于基因的重排、双链的结合，再加上拼接中的种种变异，TCR 结构多样性已达 10^{15}～10^{20}，统称为受体谱（TCR receptoire）。原始受体谱在胸腺发育中经过阳性选择（保留能以 TCR 识别自身 MHC 分子的胸腺细胞）和阴性选择（去除能以 TCR 识别自身抗原肽的胸腺细胞），使得离开胸腺进入外周血的成熟 T 细胞只能借助于 TCR 识别和自身 MHC 分子相结合的非己抗原肽。尽管后者的数量仍是很大的，但自身 MHC 分子种类却相对有限，这样，MHC 就极大地限制了 T 细胞的识别和杀伤功能。也就是说，T 细胞识别抗原要有两种识别方式，一是 TCR 与 MHC 沟槽中的特异性多肽结合，而此多肽的基序只能与某一型号 MHC 分子结合，不是与所有 MHC 分子结合；另一是 TCR 识别抗原槽两则的同种异型部位的 α 螺旋结构。由此限制了 TCR 只能识别自身 MHC 分子逞递的抗原。

小　结

免疫是机体识别和清除非自身的大分子物质、保持机体内外环境平衡的生理学反应。免疫的基本功能有抵抗感染、自身稳定和免疫监视等三大功能。免疫应答通常分体液免疫和细胞免疫。动物机体能调节免疫应答水平，以免反应过强或过弱而造成损害。所以，免疫亦有两重性，它既是重要的防御反应，亦往往陪伴着有害的病理过程，其反应程度若超越了正常的生理范围，就会导致功能紊乱和组织损伤。

由主宰或执行机体免疫功能的器官、组织、细胞和分子所组成的一个系统称为免疫系统。免疫器官分为中枢免疫器官和外周免疫器官，中枢免疫器官包

括骨髓、胸腺，鸟类还有腔上囊。在胸腺内分化成熟的淋巴细胞称为胸腺依赖性淋巴细胞，简称T细胞。法氏囊是鸟类特有的器官，来自于骨髓的干细胞，经血流进入法氏囊后，分化成为囊依赖性淋巴细胞，简称B细胞，参与体液免疫。哺乳动物并不存在独立的囊类结构，骨髓的一部分作用相当于法氏囊的功能。外周免疫器官包括脾脏、淋巴结和消化道、呼吸道及泌尿生殖道的淋巴小结。参与免疫应答的细胞统称为免疫细胞，包括T细胞、B细胞、N细胞、K细胞、单核巨噬细胞、各种粒细胞等。凡能刺激机体产生抗体和致敏淋巴细胞并能与之结合，从而发生特异性免疫反应的物质称为抗原。抗原所引起的免疫应答具有极强的针对性，一种抗原物质刺激机体产生相应的抗体，这种抗原只能与相应的抗体相结合发生反应，此即抗原的特异性。这种特异性取决于分子表面具有特殊立体构型的、有免疫活性的化学基团，这些基团易为机体的免疫应答所针对，是与抗体相结合的部位，这个区域或基团称为抗原决定簇，又称为抗原表位。抗体是在抗原和免疫系统相互作用下，由B细胞产生的能与相应抗原发生特异性结合的免疫球蛋白（Ig），通常存在于血液等多种体液中及某些细胞的表面上。根据重链稳定区的分子结构和抗原特异性不同，人类Ig可分为五类，即IgG、IgA、IgM、IgD、IgE。

抗体分子单体由四条肽链构成，两条长的肽连称为重链（H链），两条短的肽连称为轻链（L链），四条肽连的N端，即L链的1/2与H链的1/4处，称为可变区（V区），它赋予抗体以特异性，同时也是与特异性抗原结合的位点。多肽链的C端，即L链其余的1/2及H链其余的3/4部分，其氨酸数量、种类、排列顺序及含糖量都比较稳定，故称为恒定区（C区），Ig都是对称性的大分子。IgG、IgD、IgE及大多数血清型IgA均为单体，而分泌型IgA为双体，即由两个单体IgA由J链相连，再加上一个分泌片段组成；IgM为5聚体，即由5个单体Ig和1条J链相连而成。蛋白水解酶可将Ig分子裂解为不同大小的片段。

单克隆抗体是指由单个B细胞克隆生产的、在分子上是同质的抗体。其生产的原理是：能产生抗体的B细胞在体外不能持续传代，在体外能长期传代的骨髓瘤细胞不能分泌抗体。因此，将能产生抗体的免疫淋巴细胞与骨髓瘤细胞融合，经过筛选、培养、克隆化，建立既能分泌针对预防抗原的特异性单克隆抗体，又能在体外无限繁殖的杂交瘤细胞系，用来生产单抗。该项技术称为淋巴细胞杂交瘤技术。基因工程抗体是将分子免疫学关于Ig基因结构与功能的知识与DNA重组技术、蛋白质工程相结合，按人们的意愿在基因水平上对Ig分子进行切割、拼接或修饰，人工合成、重新组装的新型抗体分子。

免疫球蛋白基因由3组不连续的基因控制，即κ轻链基因群、λ轻链基因

群和H链基因群。各个基因群中含有许多基因节段，在未分化的胚系B淋巴细胞和其他有核细胞的染色体中，编码免疫球蛋白可变区（V区）和恒定区（C区）各个部分的基因节段是互相分离的。在B细胞的发育过程中，一部分基因节段互相接近，重新排列，形成功能性的免疫球蛋白基因单位，这一过程称为基因重排。完整的轻链由3个基因节段编码：V基因节段和J基因节段编码可变区，C基因节段编码恒定区。人和小鼠Igκ链可变区由多个基因节段组成。这些节段分为V_κ和J_κ链两组，κ链恒定区基因节段（Cκ）只含有单个外显子。λ基因同样也分为3组：Jλ、Vλ和Cλ基因节段。V_H基因节段编码重链可变区的疏水前导肽和N端的96～101个氨基酸，每个V_H基因节段有2个外显子。B细胞在分化成熟的过程中，从V_H或V_L的V、D和J基因节段库中各选择出一个基因节段，经DNA重排，与恒定区（首先是$C\mu$）构成一个功能性基因进行转录，经转录后加工，再成熟为免疫球蛋白。在每个V_H和V_L节段的下游、每个J_H或J_L节段的上游以及在每个D_H节段的两侧，存在着一种特殊的、相互关联的侧翼序列或称为重组信号片段（RSS）。每个侧翼序列含有一个7核苷酸的回文序列和一个9核苷酸的保守序列，两者之间有一段固定数目（12和23±1）的随机核苷酸序列分隔。在被连接的两个基因节段之间，它们是反向互补的，可使两个待连接基因节段反绕成环状结构，在DNA双链修复系统的作用下，使两个基因节段互补连接在一起。RSS的重要特点是，分隔序列数目是固定的12bp或23bp，一个侧面有12bp分隔序列的基因节段只能与一个侧面有23bp分隔序列的基因节段重组或连接，反之亦然。这一现象称为12～23bp规律。解释Ig基因节段重排或连接机理的假说均以RSS和12～23bp规律为基础。

免疫球蛋白的多样性是指免疫球蛋白的V区（V_H和V_L）的多样性，多样性产生与以下几个因素有关。（1）组合性链接，（2）接合多样性，（3）体细胞突变。抗体基因重排的控制由克隆选择学说解释。VDJ与C的连接，这个过程亦称为重链类型转换，即VDJ从与$C\mu$或Cδ连接，转换到与Cα、Cε或Cγ的连接，从而使B细胞产生具有同样抗原特异性（具有相同的VDJ）但恒定区不同（Cα、Cε或Cγ不同）、能介导不同功能的其他类型的抗体（IgG、IgA、IgE）。类型转换的机理可能是通过两种模式进行，即成环和删除模式和姐妹染色单体交换模式。

T细胞受体（TCR）是T细胞表面存在的能与抗原发生特异性结合的受体，大多数T细胞的TCR是由α、β两条链组成的异二聚体，并与CD_3一起形成完整的TCR－CD_3复合体。T细胞受体基因与Ig基因一样，TCR的可变区和恒定区基因也必须在胚系T细胞中经历基因重组才能形成功能性表达基因。TCR基因的重排及其多样性产生的机制亦与Ig基因类似。

凡能引起强而迅速排斥反应的抗原称为主要组织相容性抗原，它包括多个系列，组成复杂的抗原系统。编码这一系统的基因位于同一染色体片段，是一组紧密联锁的基因群，称为主要组织相容性复合体（MHC）。控制免疫应答能力和免疫调节功能的基因即免疫应答基因也位于 MHC 中，因此，MHC 不仅与移植排斥反应相关，而且也与免疫应答和调节有关。小鼠 MHC 称为 H2 复合体，人的 MHC 称为 HLA 复合体。

小鼠 MHC－H2 复合体位于第 17 号染色体上，可分为 3 类，其中 I 类基因包括 K 区、D 区、L 区，II 类基因为 i 区，III 类为 S 区。I 类分子由 K 区或 D、L 区编码的重链和非 H2 基因编码的 β2（12KD）所组成，主要参与移植物急性排斥、抗体产生、细胞裂解和向 CD8＋ T 细胞呈递抗原；而 II 类分子两条链分子量相近，α 链、β 链分别为 35KD 和 28KD，主要分布在 B 细胞、巨噬细胞和激活 T 细胞表面，参与移植物慢性排斥、抗体产生、移植物抗寄主反应（GVHR），激发混合淋巴细胞反应（MLR）和向 CD4＋ T 细胞呈递抗原。

人类主要组织相容性复合体（MHC）称为人类白细胞抗原（HLA），主要分成 3 组：HLA I、HLA II、HLA III，HLA I 类基因从第 6 号染色体中心侧开始依次为 HLA－B、－C、－E、－A、－G、－F 及 I 类样假基因或基因片段 HLA－17、－X、MHC I 类链相关基因 C、－30、－L（92）、－J（59）、－80、－21、－K（70）、－16、－H（54）、－90、－75、MCCD 等共 21 个基因。HLA II 类基因从中心侧开始依次为 DPB、DPA、DQB、DQA、DRB、DRA 基因。HLA III 类基因位于 II 类与 I 类基因区域之间，主要有补体 C2、C4A、C4B 基因 B 因子（Bf）基因，21 羧化酶基因（CyP 21B）CyP 21A 基因，热激蛋白（HSP70）基因，肿瘤坏死因子基因（TNF～α）TNF A、TNF B 和淋巴毒素基因（LTA），以及功能不明的 XA、RD、G11 基因。

MHC 的生物学意义有：（1）MHC 在同种异体组织移植时，若供、受体移植抗原不同，将会诱发受体产生明显的移植排斥反应。MHC II 类抗原错配启动了免疫应答，而 I 类抗原错配则导致移植物在免疫效应阶段成为被攻击的靶子；（2）MHC 分子参与胸腺细胞分化和成熟的过程。（3）MHC 对 APC 抗原起呈递作用。T 细胞在识别细胞表面抗原决定簇的同时，还必须识别细胞上的 MHC 分子，具有同一 MHC 表型的免疫细胞间才能有效地相互作用。这一现象称为 MHC 限制性。

复　习　题

1．解释下列名词：

免疫；T－细胞；B－细胞；免疫细胞；抗原；抗体；抗原性；免疫原性；

完全抗原；半抗原；抗原决定族；Ig 单体；Ig－V 区；Ig－C 区；Fab；Fc；Fc；Fd；单克隆抗体；基因工程抗体；基因重排；假基因；RSS；重链类型转换；T 细胞受体（TCR）；T 细胞分化抗原（CD）；T 细胞受体基因；主要组织相容性复合体（MHC）；免疫应答基因；MHC 的限制性。

2. 请回答下列问题：

（1）免疫的基本功能；（2）抗体分子的结构；（3）抗体分子的分类；（4）免疫球蛋白基因的结构；（5）重链基因结构的特点；（6）免疫球蛋白基因重排的顺序；（7）免疫球蛋白基因重排的控制；（8）基因节段间连接的 12～23bp规律；（9）重组信号片段的特点；（10）形成抗体多样性的原因；（11）重链类型转换的机理；（12）T 细胞基因的重排及多样性产生的机制；（13）MHC Ⅰ、Ⅱ型分子的特点；（14）MHC Ⅰ、Ⅱ基因的结构；（15）MHC 的生物学意义；（16）MHC 的限制性。

第八章

基因对发育和细胞分化的控制

高等动植物都是由许多在形态和功能上不同的细胞组成的，它们分别构成组织、器官乃至系统，然而这些不同的细胞最初都是由一个受精卵发育而来的。由一个受精卵所产生的同一个体内的各类细胞为什么会变得如此不同？即从单个受精卵长成成体的过程中，细胞的差别是如何产生的？对这些现象分子机制的研究已成为生命科学领域最活跃的分支学科之一。在这方面，研究者们应用了许多模型系统进行研究，如λ噬菌体的发育，细菌芽孢的形成，酵母不同交配型的转换，原生动物和高等真核生物发育的基因控制等等。这些研究发现了许多与发育有关的重要基因，并对基因控制发育的方式也有了较深入的了解。本章将利用几种在发育方面研究较深入的模型系统讨论基因对发育和细胞分化的控制。

第一节　基因控制发育的方式

一、发育阶段的时间决定

无论是原核还是真核生物，其生长发育过程均是一系列发育事件按程序不断展开的，展开的过程主要是由阶段性表达的发育调控基因组成的层次网络在不同发育时期、不同水平上进行控制，指导细胞的运动、分裂和分化等。因此，为了控制基因何时进行活动，细胞必须能够测量时间。事实上，真核细胞和原核细胞在细胞水平上都具有准确的细胞钟，也有人称为发育定时钟（developmental timer，DT），即各种细胞可能都有一种独立的内部钟，同时还具有一种保证在一定时间进行关键生化反应的装置（类似闹钟）。在这种情况下，细胞内事件a、b、c相互独立，事件b确定时间b，事件c确定时间c，即使事件b因某种原因不能报时，事件c仍然能够按预定计划报时。另一方面，可能不涉及某些关键反应，事件c就不能发生，除非事件b已经出现，例如事件c可能是一种化学反应，它需要事件b的产物。在这种情况下，细胞内每一事件

都依赖于前一事件，并且所有事件都以固定的顺序联系在一起。

原核生物基因表达的时序调控多是通过一种或几种蛋白质因子与 RNA 聚合酶相互作用而实现的。如枯草杆菌噬菌体 SPO1 是一种烈性噬菌体，在其生活史的早、中、晚三个时期分别由早期基因、中期基因和晚期基因进行转录。在侵染 4～5 分钟后，早期基因停止表达，中期基因开始转录，侵染后 8～12 分钟，晚期基因的转录便取代中期基因的转录。这两次转变需要三种 SPO1 自身调节基因的表达，这三种调节基因是 28、33、34。基因 28 为早期基因，寄主的 RNA 聚合酶全酶转录早期基因后就产生调节蛋白 gp28（28KD），gp28 取代 σ55 与核心酶结合，使聚合酶对于启动子的识别特性发生改变，即不再识别早期基因启动子而只识别中期基因的启动子。中期基因 33 和 34 的突变导致晚期基因不能表达，即这两个基因产物 gp33（33KD）和 gp34（34KD）又是晚期表达的调节蛋白，它们可以取代 gp28，并使 RNA 聚合酶只能识别晚期基因。这种连续更换 σ 亚基的途径使 SPO1 的早、中、晚三个阶段的基因能够有条不紊地表达。

与 SPO1 类似，大多数复杂噬菌体都使用正、负调节子来控制噬菌体侵染期间基因表达的顺序。λ 反终止蛋白 N 和 Q 正向行使功能，而 T7 的基因 α 蛋白则通过抑制细菌 RNA 聚合酶而负向行使功能。相反，σ 因子（如 SPO1、基因 28、33 和 34 以及 T4 基因 33 和 55 基因的产物）既可正向行使功能（启动晚期基因表达），也可负向行使功能（通过替换旧的 σ 因子关闭早期基因表达）。因此，即使很小的噬菌体基因组也能利用非独立事件（dependent event）的某段序列来计算时间。但是，由于 RNA 聚合酶分子沿 DNA 模板移动速度相对较低（DNA 复制比转录快 100 倍），噬菌体就能利用转录来独立测量时间。例如，在 37℃，大肠杆菌 RNA 聚合酶的移动速度为每秒 40 个 nt，所以一条聚合酶分子转录完 λ 基因组可能需要 25 分钟以上。由于几个操纵元同时进行转录，所以合成具感染力的噬菌体不需如此长的时间。

虽然我们有可靠的证据说明转录能够测量噬菌体的生物学时间，但在复杂的高等真核细胞中，研究者们只成功地解释了胚胎发生期间的时间测量问题。动物胚胎早期发育的研究进一步表明，DT 与通常的生物钟并不相同，即它不是物理学上的计时，而是一种阶段上的相对性。例如爪蟾受精卵在第十二次卵裂后从快速的、对称的细胞分裂向慢速的、非对称的细胞分裂转变；果蝇的受精卵在完成第十三次卵裂后才开始由多核体转向细胞化，这与其发育绝对时间的长短并无关系。尽管不同动物的卵裂速度不同，如蛙的受精卵在 4℃下经 43 小时可达囊胚早期，而小鼠受精卵在体内经 50～60 小时的发育才达 8 细胞期，猪的受精卵则需 66～72 小时才能达 8 细胞期，但它们都要经过相同的发育阶

段，即由受精卵开始到囊胚、原肠胚、神经胚和尾芽胚等才最后长成一个生物体。这里仅以青蛙早期胚胎发育为例，说明影响高等真核生物早期胚胎发育时间的决定因素。

两栖类动物由于其受精卵很容易获得和观察，且在实验条件下易于通过发育的正常阶段，使其很早就作为实验动物用于发育方面的研究。青蛙胚胎学的一个很重要特征是所有营养物（大多以卵黄形式贮存）必须存在于未受精的卵之中。这些营养是从囊胚到原肠胚阶段发育所必需的。只有发育的胚胎能够自己觅食时（蝌蚪阶段）才开始真正全面生长。青蛙卵相对其成体细胞而言是很大的，受精后通过细胞分裂形成的子细胞逐渐变小，这种现象叫做切割分裂(cleavage division)。

受精之后，蛙卵很快进行 12 次迅速的、几乎是同步的切割分裂，其 DNA 合成非常快，每次细胞分裂只有很短的时间。在这期间，很难测量新合成的 DNA 进行转录的情况，其原因可能是 DNA 分子上存在 RNA 聚合酶，干扰了 DNA 的迅速复制。当进行 12 次分裂之后，在受精卵发育成具有 2^{12} 个细胞（4096）的囊胚时，细胞分裂速度开始下降到正常体细胞的分裂水平，RNA 聚合酶Ⅱ和Ⅲ开始进行转录，而 RNA 聚合酶Ⅰ则稍后开始转录。这种急剧的生理变化叫做中囊胚过渡（midblastula transition)。因为这种变化发生在囊胚形成（blastulation）的中途和原肠胚形成（gastrulation）之前。囊胚细胞是如何知道它们进行了 12 次细胞分裂的呢？有两种很大可能性，第一种是囊胚直接计数每一轮细胞分裂（如细胞分裂一次后就对 DNA 上的位点按顺序进行修饰）；第二是计算细胞分裂的产物（如整个胚中 DNA 的总量)。根据实验，现在知道，在受精卵微量注射了外源 DNA 后，中囊胚过渡比正常情况下出现较早。任何外源 DNA 都有效（质粒、噬菌体或细菌 DNA)，而且微量注射的 DNA 越多，过渡出现的时间就越早。这种现象有力地证明，成熟的蛙卵含有某种预先形成的因子（可能是一种蛋白质)，这种因子以化学计量方式与 DNA 进行结合，2 倍的 DNA 与 2 倍的因子结合。未受精卵含有精确的正常含量的因子，它与 4096 个细胞中的总体 DNA 结合而被耗完时，从而启动转录和原肠胚形成所需的细胞能动性。可以看出，这种因子是一种 DNA 结合蛋白，它与任何 DNA 进行松散的、非特异性结合，从而阻止转录。最初，卵细胞中高浓度的游离因子可能会饱和所有潜在的转录单位，只有当游离的因子完全与 4096 个细胞中的 DNA 结合后，新合成的 DNA 才能恢复转录。中囊胚过渡的调节类似卵细胞特异性 5SRNA 含量的调节，即通过用 5S 基因产物稀释预先形成的转录因子 TFⅢA。稀释卵细胞中预先形成的因素对许多不同发育事件进行定时是一种精确而又恰到好处的一般性机制。但是很难设想，正在成熟的

蛙细胞或卵是如何精确合成预先形成的因子的。

上述仅讨论了 DNA 复制对 DT 作用的控制机制，即受精卵中可能存在一定量的控制因子，随着 DNA 复制的不断进行，这些因子逐渐被稀释，当其浓度降到临界点以下时，基因从抑制状态下被释放出来并开始表达。目前对细胞质因子在卵子或初期胚胎中存在的具体情况的研究也取得了一些进展。如对中胚层形成的研究发现成纤维细胞生长因子（bFGF）和活性素（activin）具有强的中胚层诱导活性，Wnt 蛋白、视黄酸、Raf－1、激酶、BMP－4 和成头蛋白（noggin）等因子在中胚层形成过程起诱导作用或重要的修饰作用；另外，细胞核/质之比与早期胚胎发育中囊胚转移期（MBT）的开始时间关系密切，从而影响到 DT 作用的启动。

二、基因控制发育的方式

基因控制发育的方式很复杂，但对所有生物一个基本的原理是一致的：即从个体发育中不同的发育事件看，基因的表达是先后有序的，发育在阶段性表达的发育调控基因推动下进行。而从整个有机体的发育看，个体发育是受多个基因群形成的相互作用的多层次网络控制的，这些基因的表达在时空上相互联系和配合，控制发育的遗传程序。目前这方面的知识主要是从对果蝇的研究得到的。大量遗传证据表明，果蝇胚胎的发育形式是由控制基因的等级系统所决定的，如控制胚胎形体（体轴和体节）可以分为两个层次：①母体基因。果蝇卵子发生期间，母体细胞中的一些基因产物（包括 mRNA 和蛋白质）进入卵子，这些基因的产物在卵质内按一定时空图式分布，从而使位于卵质不同位置的细胞核内的基因被选择性地激活，最终引起细胞分化，如果蝇正常头部发育所需的基因 bicoid（bcd）、后节发育基因 tudor（tud）、后部末端体节发育基因 torso（tor）等。②合子型基因。当卵裂到一定时候，合子型基因被激活，最先被激活的为缺口基因（gap gene），这类基因突变后能导致大范围体节缺失；缺口基因的下游基因是突变后能导致单数或双数体节消失的配对基因（pair－rule gene）；随后表达的是体节极性基因，体节极性基因的表达将胚胎再细分为 14 个条形区，其突变将导致半个体节缺失和另半个体节的反转重复；处于这一系统较低位置的是同源异型基因，它们决定各个体节的不同，其突变导致一个器官发育成另一不同器官，如 antp 突变使腿部替代触角而着生在眼周围细胞之外。上述基因层次网络中，基因相互作用的动态行为还远未弄清，但越来越多的发育调控基因在发育控制过程中被发现承担多项使命，说明发育调控所采取的基本分子作用方式有限，但组合后产生的结果却是变化多端的，在这

方面还有待于更深入研究。

三、Wnt 基因家族与发育

近年来，发育生物学研究在鉴定形态发生中起作用的信号分子方面取得了重要进展。这方面突出的例子是原癌基因在细胞分化和胚胎发育中的重要作用。原癌基因的产物包括许多分泌出来的生长因子、细胞表面受体、翻译信号的细胞质成分以及通过转录调节执行其命令的细胞核因子，这些因子是调节生物发育机制的基本成员。其中，Wnt，一个产生分泌型糖蛋白的庞大基因家族，由于在囊胚、原肠胚、中枢神经系统形成和器官发生以及肢的发育等基本的发育事件中起重要调控作用而受到了广泛重视。各种 Wnt 蛋白都是一些多功能因子，通过自分泌或旁分泌途径在不同细胞背景下诱导不同的生物学反应。Wnt 基因家族表达类型很复杂，由 Wnt 编码的蛋白的生物学活性形式及其受体分子尚有待于进一步鉴定。因此，对 Wnt 基因功能的研究一直是基因对发育调控研究中十分活跃的课题之一。这里仅对 Wnt 基因家族及其在几种动物发育调控中的作用进行介绍。

（一）Wnt 基因家族

第一个 Wnt 基因（Wnt gene）是在小鼠中寻找由原病毒插入而被激活的致癌基因时分离得到的，最初被命名为 int－1（integration 简为 int），意指 MMTV（鼠乳腺肿瘤病毒）DNA 的整合部位。由于它与果蝇中的 wg（wingless）基因具有类似的编码产物，因而将这类在不同物种间的同源基因统称为 Wnt 基因。时隔不久，与鼠 Wnt－1 同源的基因也在人和爪蟾中发现，而且还在同一物种内发现有多个同源异型基因，例如鼠除 Wnt－1 外还有 Wnt－2、Wnt－3A 和 Wnt3B 等。随后根据这些同源基因的结构保护序列设计了用于 DNA 聚合酶链式反应（PCR）的引物，利用 PCR 技术发现了 Wnt 基因家族的许多成员（表 8－1）。除表 8－1 中的物种外，目前在脊椎动物的无颌类（盲鳗）、鱼类（金枪鱼、鲨、斑马鱼）、两栖类（钝口螈）、鸟类（吐绶鸡、血雁）、爬行类（牛蛇、刺蜴）以及在无脊椎动物的线虫、蛭、海星、海胆以及酵母中也发现了 Wnt 的同源基因。由于仅用一套引物从胚胎 cDNA 文库中分离这些基因，因而 Wnt 基因家族的成员尚未全部分离出来。如果用该基因家族其他位置的保守序列设计引物和建立其他组织或器官的 cDNA 文库还可能会分离更多的 Wnt 基因家族的成员。

每一个 Wnt 基因的可读框都编码一种 350～380 个氨基酸的蛋白质，整个序列在不同物种间有 100 多个保守的氨基酸残基。一些 Wnt 蛋白还有额外的

插入序列、氨基末端或羧基末端结构域，如果蝇的 wg 蛋白质在其最后一个外显子的编码区之前有 85 个氨基酸的插入序列，DWnt－3 编码的蛋白质则有一个 155 氨基酸的插入序列和很长的氨基末端延长区，小鼠的 Wnt－2 和爪蟾的 Wnt－8 蛋白则有很短的羧基末端延长区。比较发现，多数情况下，一个物种或不同物种的 Wnt 蛋白的相同氨基酸序列在 30％～60％之间。由于此基因家族成员的高度保守性，利用小鼠不同 Wnt 基因间的同源序列可对其他脊椎动

表 8－1　几种生物中的 Wnt 基因及染色体定位

	物　　种	基　　因	染色体定位
人	*Homo sapiens*	Wnt－1	12q13
		Wnt－2	7q3.1
		Wnt－3	17q21
小鼠	*Mus musculus*	Wnt－1	15
		Wnt－2	6
		Wnt－3	11
		Wnt－3A	11
		Wnt－4	4
		Wnt－5A	14
		Wnt－5B	6
		Wnt－6	1
		Wnt－7A	6
		Wnt－7B	15
爪蟾	*Xenopus Laevis*	X－Wnt－1	?
		X－Wnt－3	?
		X－Wnt－4	?
		X－Wnt－5A	?
		X－Wnt－8	?
果腹果蝇	*Drosophila melanogaster*	wg	28A
		D－Wnt－2	45E
		D－Wnt－3	17B
线虫	*C. elegans*	Ce－Wnt－1	Ⅱ
		Ce－Wnt－2	Ⅳ

物、昆虫和棘皮动物中的同源基因进行识别。如小鼠的 Wnt－1 蛋白在爪蟾、斑马、鱼和果蝇中的相应蛋白质的同源序列均在氨基末端具有一个相同的半胱氨酸残基，而其他 Wnt 蛋白则含有两个这样的半胱氨酸残基。对 Wnt 基因的全基因结构及控制其表达的结构尚不甚了解。已克隆的一些基因多数具有 3 个内含子，位置也大致相同。但线虫（*Caenorhabditis elegans*）的内含子和外显子则有不同分布，果蝇的 DWnt－3 则无内含子存在。Wnt 基因家族与许多发育现象密切相关，如 Wnt 基因的表达失控引起青蛙体轴的过度发育，在小鼠中的过量表达则会引起乳腺癌，Wnt 基因缺失会阻止哺乳动物脑的正常发育和昆虫胚胎的正常卵裂。

（二）小鼠的 Wnt－1 基因与发育

小鼠胚胎和成年个体中基因表达的限制性特征揭示 Wnt－1 基因可能在哺乳动物发育过程的重要事件中起调节作用。随后的研究发现，Wnt 基因家族的成员在小鼠胚胎发育过程中均有相似的功能。尽管由于缺乏抗血清而不能鉴定 Wnt 蛋白在组织中的分布情况，但应用 WntRNA 进行的体外原位杂交表明，Wnt 以一种局部方式与特定细胞类型的建立有关。

研究 Wnt－1 突变体的表型进一步支持了上述结论。有人通过两种途径产生 Wnt－1 无效突变（null mutation）对此进行了研究：一种是在胚胎干细胞的 Wnt 基因的第二外显子处插入 neo，获得 Wnt－1 neo 等位基因，另一种是在 Swaying 小鼠中发现的 Wnt－1 基因可读框中间由于一个碱基缺失导致翻译终止的 Wnt－1sw 等位基因。Wnt－1 neo 和 Wnt－1sw 的纯合子或双杂合子均表现一定范围内的表型变异。一种极端类型是出生时由于缺乏小脑和部分中脑而致死（中枢神经系统的其他部分发育是完整的），在胚胎发育的 9.5 天便可检测到上述器官的缺失；另一种情况是上述突变基因的纯合子动物可活至成年，但由于小脑前半部分缺失而导致运动失调。可见，Wnt－1 基因的突变对动物最轻的损害是导致小脑前半部分不能发育，严重的则导致整个小脑不发育并影响到中脑。产生上述差异的机制尚属未知。

Wnt－1 基因在小鼠胚胎发育的 8～14 天的中枢神经系统（CNS）的表达模式与 Wnt－1 基因突变影响到的区域相符合。在胚胎发育的 9.5 天，Wnt－1RNA 就可在与后脑相连的中脑前半部分细胞中被检测到，早期表达此基因的区域发育成成年动物的中脑。在斑马和鱼胚胎发育中的研究也得到了类似结果。有人认为小脑的缺失是 Wnt－1 基因间接作用的结果，其原因是 Wnt－1 基因的突变抑制了小鼠 en 基因的表达，而 en 基因的表达对小脑后半部分的生长发育是必需的。在果蝇中的研究也得到了类似结果。

尽管正常情况下 Wnt－1 在发育早期的脑末端和脊柱的脊中线均可表达，

但 Wnt－1 突变个体的后脑和脊椎发育仍正常。这可能是其他基因尤其是 Wnt 基因家族的其他成员的表达弥补了 Wnt－1 基因造成的缺陷。Wnt－3 基因便是 Wnt－1 基因的一个重要候补者，它可在后脑末端及脊柱的顶部表达，但在中脑的后部不表达。

在成年小鼠中，Wnt－1 基因仅在雄性个体精子发生过程中正常表达。然而由于 Wnt－1 突变体 Swaying 小鼠也有受精能力，因而认为 Wnt－1 非精子发生所必需。

Wnt 基因家族其他成员的表达情况在小鼠中尚未完全清楚。研究发现，这些基因在胚胎期，在成年动物除淋巴系统以外的器官中以及在三个胚层的细胞中均有转录。这些基因在中枢神经系统发育过程中所起的作用与 Wnt－1 基因类似，但位置和作用不同。Wnt 基因还参与了肢的发育，如 Wnt－3 在外胚层的肢芽中表达，Wnt－5A 表达水平则呈梯度，即从外胚层→肢间充质→肢末端，在肢末端的表达量达最高水平，其在肢发育过程中表达的规律性与同源异型框基因（Homebox gene）类似。与 Wnt－5A 相反，Wnt－5B 也在肢芽中表达，但其表达水平不成梯度，其产物与 Wnt－5A 编码的蛋白在不同的位置起相似的作用，可能受一种相同的生化机制调控。

（三）非洲爪蟾胚胎发育中 Wnt 基因的诱导功能

爪蟾的中胚层形成是发育生物学中通过诱导进行研究的最早和最深入的生命现象。

蛙的受精卵具有极性，植物极的细胞诱导赤道区中胚层的发育，腹部的中胚层细胞将分化成肌肉和血细胞，背部的中胚层诱导神经外胚层的发育，决定胚胎的前后轴。移植实验表明，背部的中胚层细胞组成了 Speman 组织者（Speman 组织者指能够诱导外胚层形成神经组织或系统，并能和其他组织一起调整成为中轴器官的胚孔脊唇部分，可在不同的位置诱导产生额外体轴）。背部的植物极细胞分泌一种物质可引起赤道中胚层组成 Speman 组织者，并由此产生神经系统诱导因子。用 UV 射线处理，则会阻止背部中胚层以及之后的体轴的形成，最终导致腹胚产生。

分离外胚层帽可在体外诱导中胚层产生，此中胚层可进一步诱导产生神经系统，正常情况下的外胚层帽则发育成表皮组织。向囊胚中注射所需基因的 mRNA 则可阻止额外体轴的形成或挽救 UV 射线破坏的胚胎。如向小鼠受精卵中注射 Wnt－1 的 mRNA，引起体轴前部断裂，产生有两个头的胚胎；向 16－细胞的腹部囊胚中注射 Wnt－1 的 mRNA 会诱导产生完整的第二体轴。UV 损伤的胚胎几乎可以通过注射 Wnt－1 的 mRNA 完全恢复正常，说明 Wnt 基因可作为背部植物极的诱导者或称为 Nieuwkoop 中心。

用类似方法对 Wnt 基因家族的其他成员研究表明，爪蟾的 Wnt－1、Wnt－8 和小鼠的 Wnt－3A 都可诱导背部体轴的发生。上述几个基因诱导产生一种表型的现象表明，这些基因编码的蛋白质作用的受体是相同的。

（四）果蝇 wg/Wnt 基因家族

果蝇的 Wnt－1 即 wg 基因在胚胎体节极性形成方面起关键作用。对果蝇胚胎体节图式形成的广泛深入的遗传分析使这种昆虫成为研究 Wnt 基因功能的独特工具。

在胚胎发育的最初几个小时，果蝇胚胎迅速形成一个由副体节（parasegment）组成的复合体。这个过程由一系列体节形成基因控制，它们多数编码转录因子。上述结构形成后，每个副体节内部含有一套独立的体节极性基因。这些基因的突变将会导致角质层缺失或视象重复（mirror image duplication），通常表现为齿状形式发生变化。根据突变的角质层区域缺陷情况，现已鉴定出 10 多种体节极性基因。

Wg 的胚胎致死等位基因具有很强的体节极性表型，所有的无效突变都位于 Wg 的蛋白质编码区内。此基因缺失时，角质层裸露的区域为连续的齿状层所取代，分节界线也消失。Wg 无效突变（蛋白质氨基酸序列不改变）也引起中枢神经系统轻度缺陷：每个体节中只有一个二次特化的神经元。另外也有几种改变蛋白质序列的突变不导致胚胎死亡，如 wg1 是第一个被鉴别出的等位基因，它可导致器官芽（imaginal disc）尤其是翅盘（wing disc）的发育失常。最初认为 wg 基因可能是一个信号分子，但根据对 wg 突变体细胞的克隆分析，表明嵌合体中的表型是非自主性的，如 Wg 突变引起的缺陷可被周围多种野生型的细胞所修复。

早期胚胎中，wg 仅在准体节边缘前部的 1～2 排细胞中表达，其表达由较早启动的体节基因启动，如配对规律基因（fushi tarazu）和平跳基因（even－skipped），随后与其他体节极性基因相互依赖，以一种非常复杂的方式进行互作。整个胚胎中，Wg 基因 RNA 和蛋白质的染色带型几乎是相同的，这些蛋白质附在繁殖细胞上或与胞外基质相连，其行为类似哺乳动物中的相应成分，也与水蛭中 Wnt 基因产物的作用方式类似，也类似爪蟾卵中注射 wg RNA 的细胞中合成的蛋白。在电镜下可观察到部分 wg 蛋白质，产生该 RNA 细胞的临近细胞对其有胞吞作用，表明这些蛋白质分布在细胞间。

Wg 影响果蝇胚胎内许多其他基因表达，这种调节互作最为熟知的例子是扩展的种质带（germ band）胚胎的外皮层（epiderms）中 wg 对 en 的作用（图 8－1）。en 编码一种同源异型的转录因子，在每个体节的后部表达，其表达的细胞带恰好位于 wg 基因表达的细胞的下一行细胞。wg 基因突变引起 en RNA

及其蛋白质的含量下降，相反，en是保持wg表达所必需的。现在认为，调节控制wg的细胞间信号是hgh（hedgehog）基因的产物，这种信号可能通过阻止ptc（patched）细胞表面蛋白抑制wg基因表达来激活wg基因。

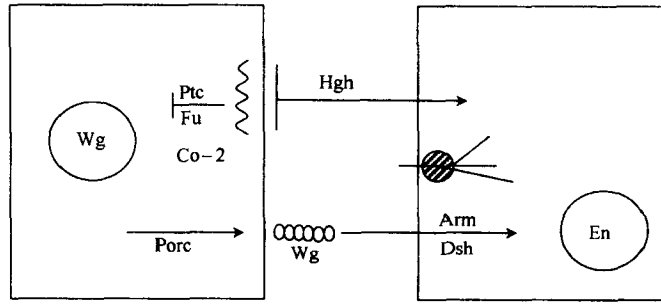


图8-1 wg和其他体节极性基因产物间互作的简图

Wg对邻近细胞中en基因的表达进行正调节，porc基因的产物（一个非自主体节极性基因）是Wg行使功能所需的，Wg受体尚不清楚；Arm和Dsh蛋白是wg和en之间的调节途径所需的；Arm蛋白与β连环蛋白（一类细胞骨架蛋白）具同源性，因此不能与细胞骨架和转膜蛋白形成复合体；Ptc蛋白对Wg的转录进行负调节；Ptc是一种复合转膜蛋白，通过Fu（一种丝氨酸/苏氨酸激酶）和前缘蛋白质的作用对wg的转录进行负调节；Hh可抑制Ptc

Wg还通过自分泌或旁分泌途径控制arm（armadillo）基因产物的定位。arm基因也是wg类中的一种体节极性基因，编码与脊椎动物斑珠蛋白（plakoglobin）和β连环蛋白类似的一种蛋白质，其RNA和蛋白质存于与胚胎的所有细胞中，这些蛋白质与F-肌动蛋白质同位于细胞内膜表面。

尽管对wg基因和其他分节基因的研究已很深入，但wg基因的受体仍未分离到。arm和dsh基因可能行使wg受体的某些功能，因为其表达需要wg的正常表达。

目前对wnt基因的了解还很肤浅。仅知道它们在细胞间信号传导和癌变中起重要作用。除wg外，其他基因的表达情况尚未完全揭示，尤其是Wnt的受体及其信号传递的途径也属未知。

第二节 λ噬菌体的发育调节

一、λ噬菌体生长概况

如第六章所述，λ噬菌体是以大肠杆菌为宿主的一种温和噬菌体，具有两

种不同发育途径。第一种是进行裂解生长，即噬菌体感染大肠杆菌后，不整合到细菌染色体上，直接利用细胞内的酶和代谢原料进行自我复制，并合成噬菌体基因的各种产物，如酶和外壳蛋白，组装出许多子代的噬菌体颗粒，最终使细菌裂解死亡，并释放出子代噬菌体。在37℃环境中，这样一个过程约需40～45分钟，每一个大肠杆菌一次可产生出100个有侵染活性的杂合子噬菌体颗粒。λ噬菌体的第二种发育途径是溶原生长，即它感染大肠杆菌后，可将其基因组整合到细菌染色体上，并成为细菌基因组的一部分，随着细菌染色体DNA的复制而复制。此时，λ噬菌体看起来只是对寄主无害的共生病毒，和寄主一起生长繁育与死亡。这种整合有噬菌体基因组的细菌称为溶原细菌，溶原细菌受某些外界物理或化学因素（如紫外线，丝裂霉素C等）的作用，原噬菌体可以脱离细菌染色体而进行自我复制，最终导致细菌裂解，游离出大量噬菌体。λ噬菌体的两种发育方向是基因组中不同基因表达的结果，但是无论最后结果如何，λ噬菌体感染细胞过程中的第一个生长阶段都相同。线状λ基因组的两端都具有长12个碱基的能够互补的单链尾巴，当噬菌体将其DNA注入到宿主细胞后1分钟内，线状DNA两端的粘性末端通过碱基配对连接在一起，形成共价结合的环状分子（图8－2）。λ噬菌体基因组中，功能上有关的基因前后连接在一起。从A到F基因与噬菌体头部形成有关，从Z到J基因与尾部形成有关。在其他一些基因中，int和xis基因编码与λ噬菌体基因组的整合和切割有关的蛋白质，基因exo和bet.与重组有关，CⅠ、Ⅱ、Ⅲ、Ⅳ基因（CⅠ、Ⅱ、Ⅲ、Ⅳgene）、N基因（N gene）和cro基因（cro gene）编码的蛋白质调节λ噬菌体其他基因表达。基因O和P编码与启始λDNA复制有关的蛋白质，基因Q的编码产物激活所有与头、尾形成有关的基因表达，而基因S、R和RZ则与宿主细胞裂解有关。λ噬菌体的两种发育途径都是一系列基因有控制地进行表达的结果。

当λDNA进入缩主细胞后，首先由启动子P_L和P_R启始编码抑制子的基因转录（图8－3）。左方操纵元的第一个基因N基因编码一种转录反终止因子，使RNA合成继续朝N基因左向进行和使cro基因的转录朝右向进行，从而转录完其余早期基因。由P_R控制的右方早期基因中，Q基因的编码产物启始晚期蛋白基因表达即裂解途径处于主导地位。启动子P_R控制的转录过程继续一段时间后，细胞才能累积足量的Q蛋白，以启始晚期基因转录。

早期基因cro和CⅡ对确定噬菌体的发育方向极为重要。在噬菌体侵染的早期，宿主细胞中累积有Cro和CⅡ蛋白，但这两种蛋白质的作用相反。CⅡ蛋白活化从一个特殊启动子P_{RM}开始的转录，合成大量噬菌体抑制蛋白即CⅠ

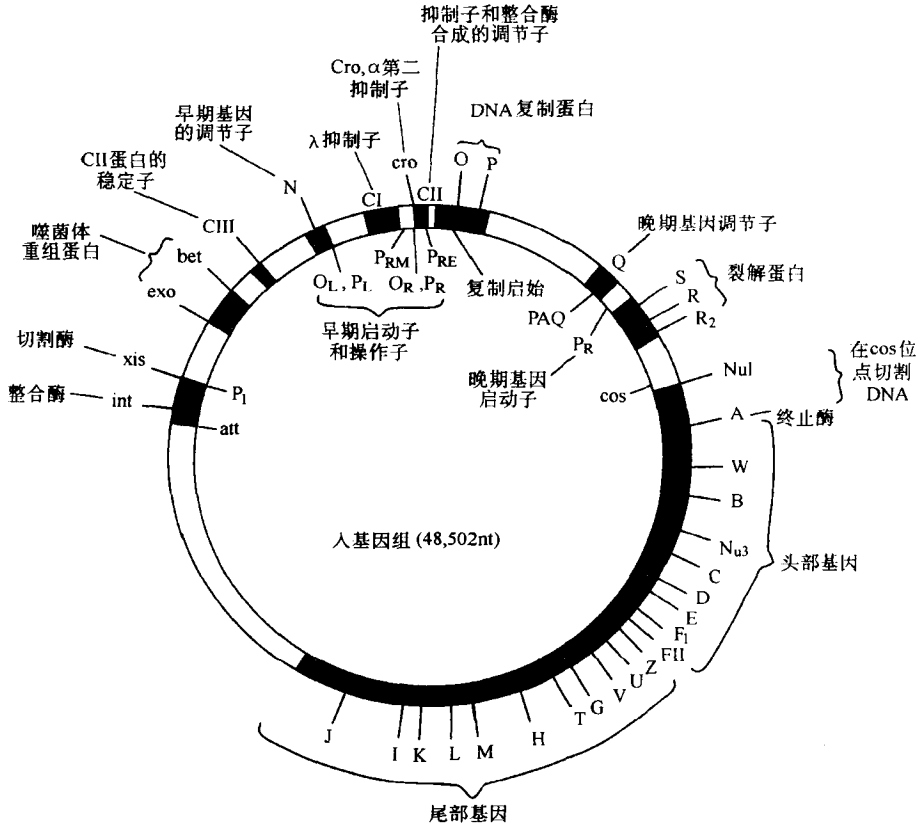


图 8－2 细胞内环型λ噬菌体图

示主要的λ基因。噬菌体 DNA 在包裹之前，在 cos 位点切割 DNA，形成线状分子，其一端在 A 基因附近，另一端在 R 基因附近

蛋白，从而使发育朝溶原性生长进行。当合成足量的CⅠ蛋白后，这种抑制子关闭早期启动子，从而抑制裂解生长的所有噬菌体基因。特别是其中的 Q 基因不能表达，从而导致不能合成裂解蛋白和噬菌体外壳蛋白。另一方面，Cro 蛋白指导噬菌体的裂解生长，这种蛋白质只降低但并不关闭由 P_L 和 P_R 启始的转录，从而降低CⅡ蛋白的CⅠ抑制蛋白的合成水平。Cro 蛋白还降低 O、P 和 Q 蛋白的合成，但能使这些蛋白质的合成达到足以完成裂解生长的水平。Cro 的第二种重要作用是通过阻止启动子 P_{RM} 来抑制CⅠ抑制子的合成。在溶原性生长期间，启动子 P_{RM} 启始CⅠ抑制子的合成。所以，Cro 蛋白反向作用于抑制基因CⅠ和抑制刺激抑制子合成的CⅡ基因的表达。

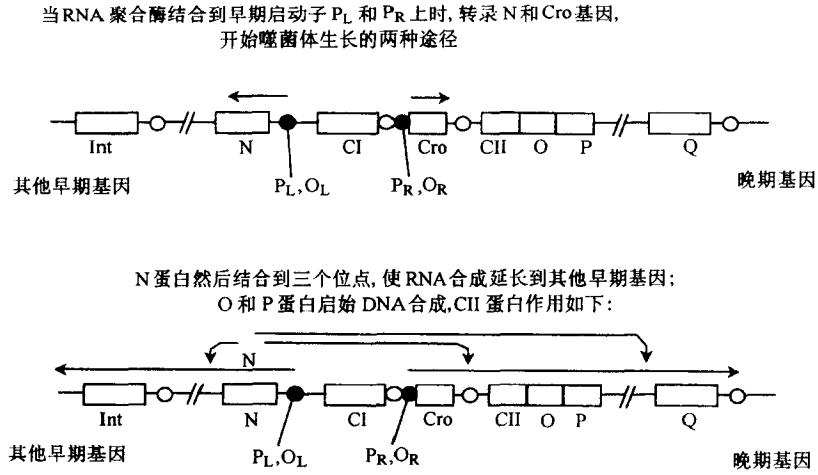


图 8－3　噬菌体λ沿两种不同途径发育期间基因表达情况

根据以上所述，λ噬菌体是进入裂解途径还是进入溶原性途径主要取决于Cro蛋白和CⅠ蛋白。在任何一个受感染的大肠杆菌细胞中都只能有一种发育途径处于主导地位，在细菌生长的正常条件下，λ噬菌体进行裂解生长还是进行溶原性生长的机会是均等的。

二、CⅠ抑制子和Cro蛋白的作用机制

Cro蛋白和CⅠ基因的产物都是抑制子，由于历史的原因，只有CⅠ基因产物叫做抑制子。许多抑制子都是二聚体蛋白，都与操纵元中操作子结合。CⅠ抑制子和Cro蛋白也同样是二聚体，二者都可能与操作子O_L和O_R结合，控制由P_L和P_R启始的RNA合成过程。操作子O_L和O_R的结构和功能都非常相似，二者都由三个相邻的亚操作子组成，即CⅠ基因左侧的O_{L1}、O_{L2}和O_{L3}，右侧的O_{R1}、O_{R2}和O_{R3}。抑制子与O_{L1}结合就足以抑制P_L启动子。但抑制子和Cro与右侧位点结合时，情况就比较复杂。在O_R中，抑制子和Cro蛋白都可通过与O_{R1}、O_{R2}和O_{R3}结合来控制2个重叠启动子P_R和P_{RM}的活性（图8－4）。

每一个亚操作子都有2段对称结构，所以总共有6段相似但不完全相同的序列，其共有序列为：TATCACCG。这些结合位点的微小差别会造成它们与抑制子和Cro蛋白的结合强度不同，从而使抑制子和Cro蛋白从事不同分工。如果抑制子和Cro蛋白的浓度都很低，那么只有结合最紧的位点才含有蛋白质，其他位点都是空的。另外，DNA结合位点的序列差别还对抑制子结合到O_R上产生相反的作用。

如果抑制子和Cro蛋白都不存在，RNA聚合酶则与P_R结合，转录右向的基因。在这种条件下，RNA聚合酶并不与启动子P_{RM}结合。如果抑制子比较丰富，抑制子就结合到O_{R1}和O_{R2}上，从而阻止P_R行使功能（图8－4）。在O_{R2}上的抑制子二聚体相连，这种抑制子二聚体之间的互作叫做协调性。协调性能够促进抑制子有效地行使抑制功能。

抑制子结合到O_{R2}上除了具有抑制启动子O_R的作用外，还具有活化启动子P_{RM}的作用，即抑制子对RNA聚合酶与P_{RM}结合起着活化子的作用。P_{RM}是控制抑制子自身mRNA合成的启动子，所以抑制子具有促进自身合成的作用。

最后，当抑制子浓度很高时，抑制子就与O_{R3}结合，抑制P_{RM}，从而防止自身的过量合成。因此λ噬菌体抑制子的合成是从正、负两个方向进行自我调节的。对于Cro蛋白而言，这种抑制蛋白与3个O_R亚操作子中O_{R3}的结合能力最强，结合最紧，所以它的第一效应是阻止P_{RM}活动，阻止形成溶原性所必需的抑制子合成。当Cro蛋白的浓度较高时，也与O_{R1}和O_{R2}结合，从而活化

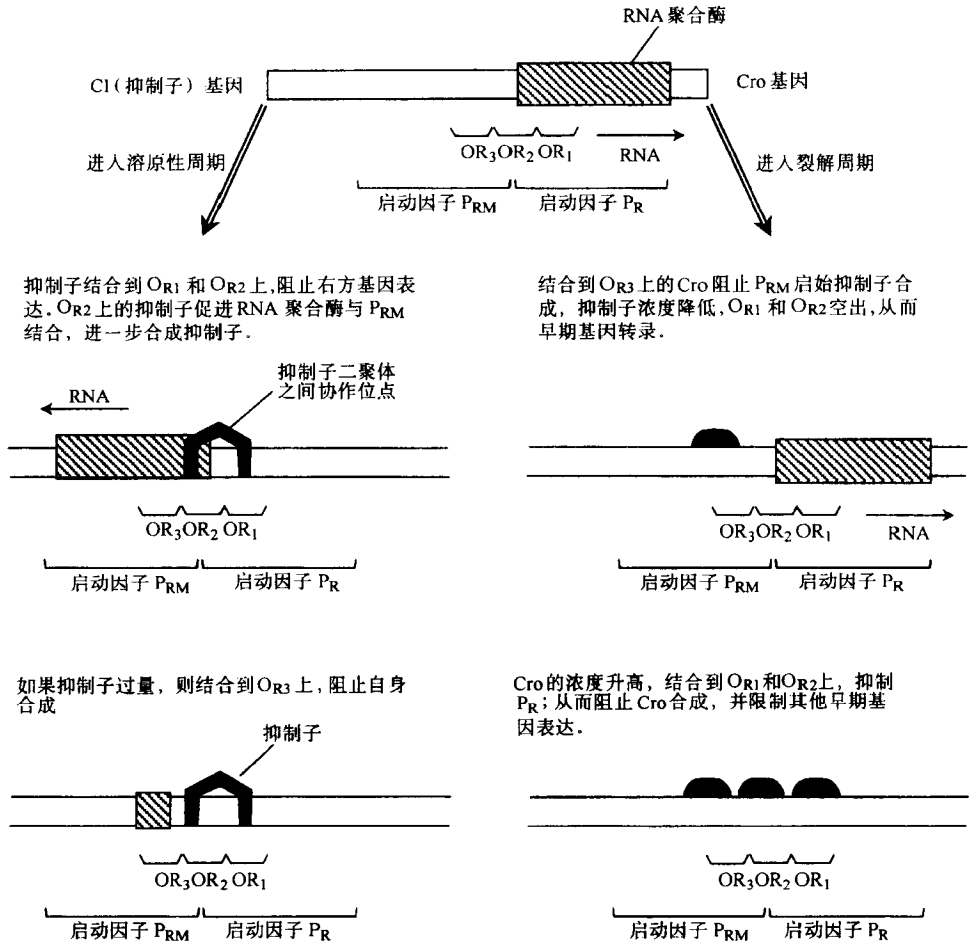


图 8－4　在λ噬菌体两种发育途径的各个阶段上，抑制子、cro 和 RNA 聚合酶占领右方操作子、启动子 P_R 和 P_{RM}的模型

右方启动子，启始右方一些重要基因表达。

三、CⅡ蛋白与噬菌体发育

在一个受感染的细菌细胞中，抑制子和 cro 蛋白中只有一个起作用。显然，这两种蛋白质在含量方面的平衡性至关重要。CⅡ蛋白是使噬菌体向溶原性发育的一种重要蛋白质，能够诱发抑制子突然合成。CⅡ蛋白为一种活化蛋白，它可以诱导 RNA 聚合酶在另一启动子 P_{RE}位置启始 CⅠ基因转录，合成

抑制子。启动子位于 cro 和 CⅡ基因之间，控制左向的转录过程。抑制子在合成之后便结合到 O_{R1} 和 O_{R2} 上，从而活化 P_{RM}。P_{RM} 的功能是在溶原性生长期间转录 CⅠ基因，合成抑制子。CⅡ蛋白的基本功能是溶原性生长所必需的，因为噬菌体的 CⅡ突变体只进行裂解生长。

CⅡ蛋白的活性受细菌生长的某些环境条件的影响，当细菌生长的环境条件改变时，其活性也会发生改变。例如，当葡萄糖充足，而 cAMP 含量很低时，CⅡ的活性也很低，所以 λ 噬菌体就从溶原性生长转入裂解生长。在细菌细胞中，CⅡ蛋白很不稳定，其浓度和总体活性都受细胞内水解蛋白酶的制约。由于 CⅡ蛋白的水平受上述环境因素的控制，可能它是细胞生长的一种传感因子，使噬菌体作出进行何种生长途径的决定。

上述 P_{RE} 启动子通常是不活动的，只有存在 CⅡ蛋白时才有活性。启动子 P_{RE} 的 −10 位和 −35 框与细菌启动子的同感序列差别很大。CⅡ蛋白在 −35 位结合到 DNA 的一侧，而 RNA 聚合酶则结合到另一侧。

第二个依赖于 CⅡ蛋白的启动子为 P_I，其 DNA 序列与 P_{RE} 相似，位于整合酶基因 int 之前，控制 int 基因转录。整合酶催化 λDNA 整合到细菌染色体上，形成原噬菌体。

第三个依赖于 CⅡ蛋白的启动子为 P_{RQ}，该启动子位于 Q 基因中间，其功能是推迟裂解发育，促进溶原性发育。

四、整合酶基因表达的逆调节

由于 CⅡ蛋白活化整合酶基因（int gene）的启动子 P_I 抑制子基因的启动子 P_{RE}，因此，一切有利于 CⅡ蛋白活性的条件都会导致突然合成抑制子和整合酶。但是 int 基因转录可以由两个不同启动子控制，即可以从 P_L 转录，也可以从 P_I 转录。所以可以认为，即使不存在 CⅡ蛋白，整合酶也能合成。但事实并非如此，因为由 P_L 启始转录的 int mRNA 被宿主细胞的核酸酶降解，而由 P_I 启始转录的 mRNA 则很稳定，能够翻译成整合酶。

分别由 P_L 和 P_I 启始转录的 int mRNA 的 3′部分的结构不同。由 P_I 启始的 RNA 合成由 int 基因之后约 300bp 位置的一个终止子所终止，这个终止子具有一个典型的茎环结构，而且其后还接有 6 个碱基 U。另一方面，当启动子 P_L 启始 RNA 合成时，RNA 聚合酶受到转录反终止子（antiterminator）N 蛋白的修饰，从而使转录通过并超出终止子。然后转录子形成一个 RNA 主杆并被核酸酶 RNA 酶Ⅲ切割，为外切核酸酶提供启始 mRNA 降解的末端，以破坏 mRNA。由于这种与负调节器有关的位点位于它作用的基因的下游，并且降解

过程从该位点开始向回进行，所以就将这一过程称为逆调节（retroregulation）。

目前对逆调节的生物学功能了解比较清楚。当CⅡ蛋白的活性降低，并且噬菌体倾向裂解发育时，由于不需要整合酶的作用，所以其mRNA就受到破坏。当CⅡ的活性很高，噬菌体倾向溶原性发育时，int基因进行表达，促进噬菌体DNA整合到细菌染色体上。

另一方面，当原噬菌体受诱导时，开始合成整合酶和切割酶，再次催化形成游离的噬菌体DNA，即通过重组，噬菌体DNA脱离细菌染色体。由于噬菌体附着位点（重组位点）P位于int基因一端至编码延长的mRNA主序列之间，所以需要合成稳定的整合酶mRNA，因此，须从int基因的一端将引起降解的位点除掉，PL启动子才能稳定地启始int mRNA的转录，合成足量的整合酶来催化切割过程。

五、N和Q蛋白的反终止作用对λ噬菌体发育的正控制

（一）受N和Q蛋白调节的操纵元

当λ噬菌体感染大肠杆菌后，N蛋白在多个位点上阻止两个早期操纵元（图8-5）的转录终止过程，使大肠杆菌的RNA聚合酶更有效地转录O基因和P基因以及右方操纵元更下游的Q基因。而Q基因的产物Q蛋白又阻止右方晚期操纵元的转录终止过程，使噬菌体形态建成和裂解细胞有关的晚期基因更有效地表达。因此就把λ噬菌体的N和Q蛋白叫做反终止因子。N蛋白作用的位点叫做nut位点，分别位于两个早期操纵元的启动子与第一个终止子之间。Q蛋白的作用位点叫做qut位点，位于受其调节的晚期操纵元的启动子之中。虽然某些λ噬菌体的终止子（包括tR1和tR2）都需要大肠杆菌的ρ因子以终止转录过程，但对于N和Q蛋白的作用来说，无论终止过程是否依赖于ρ因子，这两种反终止蛋白都能在nut位点或qut位点下游的任何终止子上阻止转录终止过程。

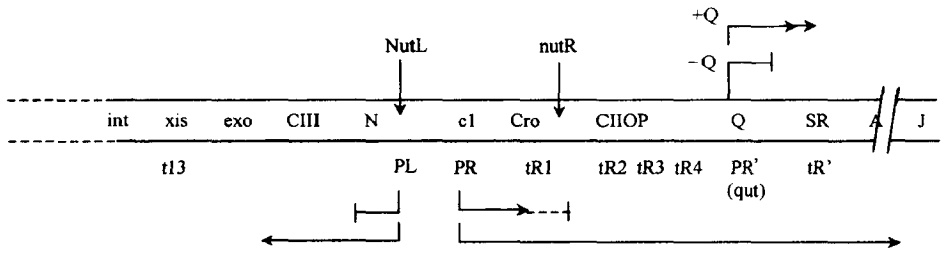


图8-5　N蛋白和Q蛋白的识别位点及其对转录的作用

（二）N蛋白的反终止作用

关于N蛋白反终止作用的模型如图8－6所示。这种调节机制的重要特征就是反终止控制信号位于新合成的RNA之内，并通过RNA环化和多蛋白质－蛋白质互作与RNA聚合酶结合在一起。当RNA聚合酶转录反终止控制信号下游的DNA时，这些转录延长因子就结合到RNA聚合酶上。

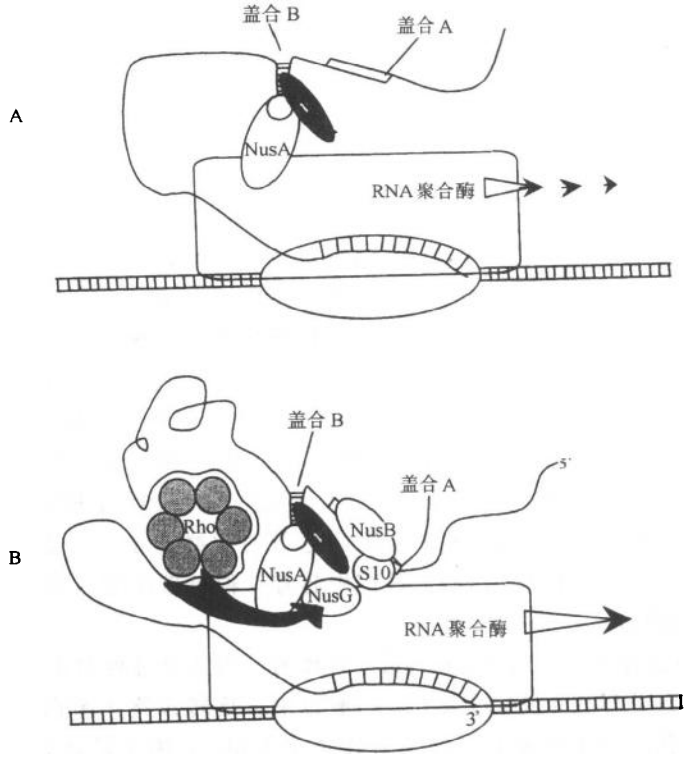


图8－6　N蛋白的反终止作用与RNA环化和多蛋白质互作有关

A. 非渐进反终止作用，N蛋白和NusA蛋白与RNA聚合酶形成不稳定复合体，其中N蛋白阻止终止过程。B. 渐进反终止作用，N蛋白、NusA、B.G和核糖体蛋白S10与RNA聚合酶形成稳定复合体，其中NusG蛋白与ρ因子互作，阻止其从新合成的RNA上释放下来，从而阻止转录终止过程

1. 转录延长复合体的组装　在宿主大肠杆菌细胞中存在4种调节转录终止和反终止的蛋白质，分别叫做NusA、NusB、NusE（核糖体蛋白S10）和NusG，这4种蛋白质都有由宿主基因编码。如果终止子比较接近nut位点，无论转录过程是否依赖于ρ因子，NusA和N蛋白能够在nutL和nutR位点下游相距5～10Kb的位点上阻止转录终止过程。在离体条件下，在含有NusA、NusB、NusG和S10蛋白质的反应液中，N蛋白具有反终止作用。N蛋白同这

4 种宿主因子一起牢固地结合到正在延长的转录复合体上，但稳定复合体的形成需要模板 DNA 中的 nut 位点（图 8－6A）。在离体条件下，λ 噬菌体的转录反终止作用需要上述 5 种因子。

N 蛋白诱导反终止复合体的形成依赖于转录复合体内多蛋白质的互作。N 蛋白和 NusA 蛋白之间的互作以及 NuaA 与 RNA 聚合酶之间的互作是这两种蛋白质形成反终止系统的基础（图 8－6A）。但是，如果没有 nut 位点和其他因子，这种结合过程是不稳定的。在稳定复合体中，N 蛋白和 NusB 蛋白通过与 NusA 和 S10 互作而间接地与 RNA 聚合酶结合。在这种受 N 蛋白修饰的转录复合体中，各种蛋白质－蛋白质互作都很弱，所以各种蛋白质之间必须高度协调才能组装反终止转录复合体。

2．Nut 位点由 RNA 组成　N 蛋白（N protein）的反终止作用所需的 nut 位点含有 2 个亚结构即 A 框和 B 框（图 8－6B）。N 蛋对 B 框具有特异性，所以它是 N 蛋白的识别位点。具有功能的 nut 位点存在于 RNA 中而不是位于 DNA 中。Nut 位点中的 2 个亚结构都与某些特殊的延长因子互作，N 蛋白只与其中 B 框结合，而 NusA 则与 N 蛋白－nut 位点复合体结合。在离体条件下，这种由三种成分组成的复合体能够保护 B 框免遭核酸酶降解。

3．RNA 环化　RNA 聚合酶是如何在反终止序列的远下游保持反终止构型的呢？只有一种可能性，就是 RNA 环首先将 nut 位点的 RNA 连接到 RNA 聚合酶上（图 8－6），当 RNA 聚合酶到达下游的终止子时，仍与 nut 位点相连。所以 RNA 只是在复合体形成过程中起着一种纽带作用，使 N 蛋白保留在 RNA 聚合酶附近。

4．反终止作用（antigermination）的机制　当转录过程中止时，RNA 聚合酶停止前进，并释放 RNA 转录子。一般说来，反转录终止蛋白能够抑制其中任何一个步骤。对于依赖于 ρ 因子的终止子来说，ρ 因子沿新合成的 RNA 运动，最后在 RNA 聚合酶的活性位点上将 DNA－RNA 杂种分子解螺旋。转录是否终止可能依赖于 ρ 因子是否追得上前进的 RNA 聚合酶。因此 N 蛋白可能通过提高 RNA 聚合酶的前进速度或防止其停滞在 tR1 终止子上来组织终止过程。N 蛋白也可能抑制依赖于 ρ 因子的转录子的释放过程。在诱导转录终止时，NusG 在 RNA 聚合酶和 ρ 因子之间起着一种桥梁作用。但是，ρ－NusG 互作也可能对 N 蛋白的反终止功能起着重要作用。因此 N 蛋白不能通过稳定 NusG 蛋白转录复合体之间的结合过程而阻止终止过程。NusG 与 ρ 因子之间进行互作，阻止 ρ 因子到达 RNA 聚合酶活性位点上 RNA－DNA 杂种分子（图 8－6b）。相反，在未修饰的转录复合体中，松散结合的 NusG 分子在某种程度上帮助 ρ 因子终止转录过程。

（三）Q蛋白的反终止作用

1. Q蛋白阻止RNA聚合酶停滞　在不存在Q蛋白（Q protein）的条件下，晚期启动子P_R启始转录后，RNA聚合酶在距转录启始位点下游16～17个碱基的位置停留数分钟（图8－7），然后离开这一位点，在终止子tR′上终止转录过程。Q蛋白能够识别滞留的转录复合体，并以此作为基质加速RNA聚合酶离开滞留位点和使其通过tR′，到达晚期基因区。无论转录过程是否依赖于ρ因子，Q蛋白能够抑制所有终止子上的转录终止过程。Q蛋白行使功能也需要宿主因子，但没有N蛋白所需的宿主因子复杂。Q蛋白能够作用于滞留在+16位的聚合酶复合体，但加入NusA结合蛋白可促进其反终止作用。

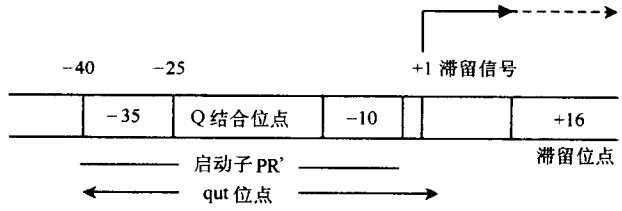


图8－7　反终止因子与反终止子的互作

Q蛋白识别晚期基因启动子$P_{R'}$，转录起点上游的qut位点与$P_{R'}$重叠；qut位点+1下游的序列使RNA聚合酶滞留其上，并发生构型变化，便于Q蛋白对RNA聚合酶进行修饰

2. Q蛋白为一种DNA结合蛋白　Q蛋白的作用位点（qut位点）位于晚期启动子$P_{R'}$的－26至+18序列之中。qut位点+1位下游的区域为一种滞留信号，因为在这一区域中发生突变会抑制Q蛋白的反终止作用。$P_{R'}$启动子的－13至－15之间的序列中发生突变会阻止Q蛋白与－10至－30之间的DNA序列结合，从而抑制反终止作用。在不存在RNA聚合酶的条件下，虽然Q蛋白能够与这一段序列结合，但是当转录复合体滞留在+16位时，Q蛋白也只与同一区域结合。如果存在NusQ蛋白，Q蛋白就能诱导RNA聚合酶与启动子之间的接触部位发生变化。可能这种变化对Q蛋白行使功能极为重要，而且在RNA链延长期间可能一直保持这种变化。

3. Q蛋白修饰RNA聚合酶的机制　Q蛋白在启动子$P_{R'}$远下游的位点上阻止终止过程的机制目前还不清楚。对于与RNA结合实际的反终止蛋白如N蛋白来说，控制信号和反终止蛋白与转录复合体结合可以通过RNA环化而保持下去。但对于与DNA结合实际的反终止蛋白如Q蛋白而言，可能需要形成DNA环（图8－8A），但是一条与qut位点结合的Q蛋白分子只能与一条转录晚期操纵元的RNA聚合酶分子接触。

另一种可能性是，当Q蛋白离开+1的滞留位点时，它从DNA上被释放

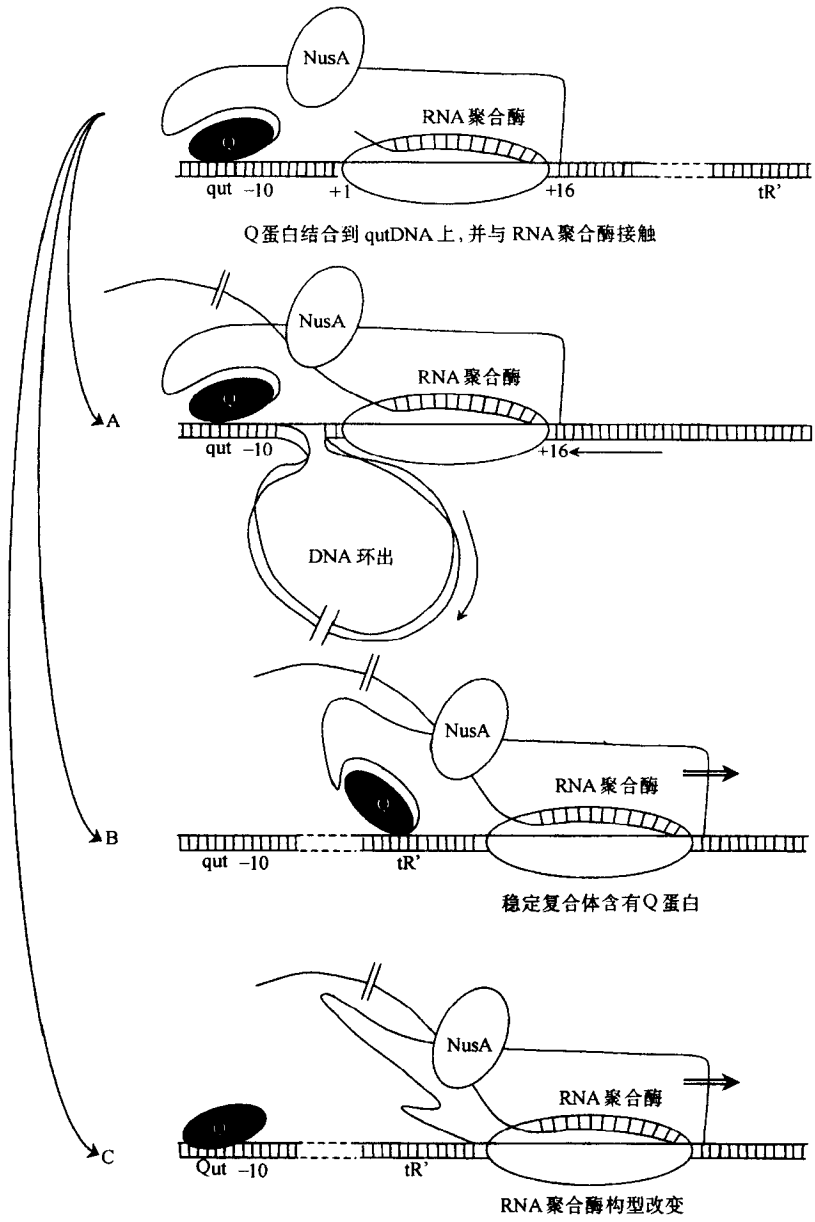


图 8－8　Q 蛋白反终止模型

Q 蛋白首先与 qut 位点的 DNA 序列结合并与滞留在 +16 位的 RNA 聚合酶互作。A：DNA 环化，使 Q 蛋白同时与 RNA 聚合酶和 qut 位点接触。B：Q 蛋白从 qut 位点上释放下来，同 RNA 聚合酶牢固结合并随其一道运动。C：Q 蛋白修饰 RNA 聚合酶的构型或结构，然后 RNA 聚合酶脱离 Q 蛋白继续转录

下来并随 RNA 聚合酶一起运动（图 8－8B）。由于缺少 qut 位点时 Q 蛋白就不能影响 RNA 聚合酶，可能 qut 位点或 Q 蛋白质在 +16 位诱导 RNA 聚合酶形成特殊构型，从而迫使 Q 蛋白质在 RNA 链延长期间保持与 RNA 聚合酶结合（图 8－8C）。

六、噬菌体 DNA 复制的特殊启动因子

λ噬菌体的 O 和 P 基因编码复制启动因子，这两种蛋白质指导细菌的 DNA 复制体结合到噬菌体 DNA 的复制启始区上。由于 O 蛋白能够识别 λ 的复制启始区，这种因子相当于大肠杆菌的 dnaA 蛋白，所以 λDNA 在宿主细胞内复制时并不需要 dnaA 蛋白。O 蛋白也能与 P 蛋白结合，这种复合体然后再与大肠杆菌的复制蛋白 dnaB 互作，因此 P 蛋白的作用是将 λ 噬菌体 DNA 连接到细菌的复制酶上。一般认为，λ 噬菌体 DNA 复制时除了不需要 dnaA 和 dnaC 蛋白以外，必须利用大肠杆菌的大多数复制蛋白。λ 噬菌体的复制启始区位于 O 基因之内，编织到 O 蛋白的编码序列之中（图 8－9）。复制启始区含有 4 个相邻的、长 20bp 的重复序列，每一重复都含有两个对称的有关的次级位点。每一重复都与一个 O 蛋白二聚体结合。O 复制启始区还含有一段富含 AT 的序列和两段回纹对称结构。根据电镜观察，O 蛋白复制启始区复合体为一种环状结构。

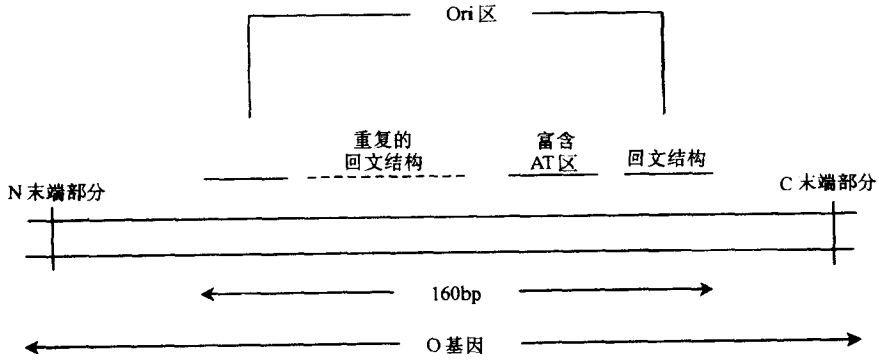


图 8－9 λ噬菌体在 O 基因中 ori 区的结构和位置

当 O 蛋白与 λDNA 的复制启始区结合后，便开始 DNA 复制过程。但是，即使存在 O 和 P 蛋白，也只有 RNA 聚合酶转录了复制启始区或其附近的 DNA 序列以后，DNA 复制才能开始。一般认为，转录使 DNA 溶解，将模板链暴露出来，以便复制启始复合体能够指导引物酶结合到模板上，合成引物 RNA。复制启始区中有一段富含碱基 AT 的序列，位于 O 蛋白结合位点的右

侧，这一区域可能与复制启始区的溶解有关。λDNA 在离体条件下复制时，需要 O 和 P 蛋白、RNA 聚合酶以及其他宿主细胞的复制蛋白。

七、单位 λ 基因组的形成

在 λ 噬菌体的生长循环期间，其 DNA 复制特征会发生变化。在侵染早期，具有感染力的环状 DNA 从 O 基因中的复制启始区开始，进行双向复制，形成多个环状分子。然后，复制方式发生转变，进行滚环复制，产生一长串具有多个基因组单位的 DNA 分子。另一方面，在滚环滚动的同时，引物酶连续地启始后随链的合成，所以复制产物为线状双链 DNA 分子。但是，在存在宿主细胞的 recBC 酶的条件下，双向复制则不能转变成滚环复制。为了防止宿主细胞阻碍复制方式的转变过程，λ 噬菌体则以其自身基因编码的 r 蛋白来抑制 recBC 酶的作用。λDNA 在进行包裹时，由噬菌体基因 A 和 nul 基因编码的终止酶将这种基因组串联体切割，并将切割下来的 λ 单位基因组包裹到噬菌体头部之中，因为环状 λDNA 不适合包装。

八、λ 噬菌体裂解周期的诱导

虽然 λ 噬菌体的溶原性周期相当稳定，但并不是不可逆转。在一个溶原性细胞中，如果 λ 抑制子突然失活，λ 原噬菌体就会进行裂解生长。原噬菌体是通过整合过程的可逆反应脱离宿主 DNA 的，这一过程由整合酶和切割酶催化。某些轻微方式如温度敏感性的抑制子的热失活等也能够诱导原噬菌体进行裂解生长。大肠杆菌 DNA 因损坏而形成的单链 DNA 在与 recA 蛋白结合后，使 recA 蛋白具有水解酶功能，破坏 SOS 基因的抑制因子 lexA 蛋白，而且活化的 recA 蛋白，还能切割 λ 抑制子。此外，用损坏 DNA 的因素如紫外光、丝裂霉素 C 等都可以诱导原噬菌体进入裂解生长周期，从而裂解细胞。通常认为，当这些诱导因素使细胞受到损坏而威胁到原噬菌体时，噬菌体进行裂解生长是一种逃脱危险环境的方式。

第三节　细菌发育的基因控制

一、细菌芽孢形成的基因控制

细菌在不利条件下会产生分化状态的细胞以维持存活，芽孢形成（spora-

tion）便是其中一种。芽孢形成是一种原始的细胞分化，从形态上看可以分为七个阶段 0－Ⅶ（图 8－10）：阶段 0 是营养体；阶段Ⅰ是两个核质体联成一个；阶段Ⅱ的特点是两个核质体中间出现一个膜；阶段Ⅲ的细胞中出现脱离的芽孢前体；阶段Ⅳ细胞中的芽孢前体外面出现胚臂和皮质层，同时在显微镜下呈折光；阶段Ⅴ的细菌芽孢中外面出现外壳，同时出现辛醇的抗性；阶段Ⅵ中芽孢成熟；阶段Ⅶ中细胞解体并释放出成熟芽孢。

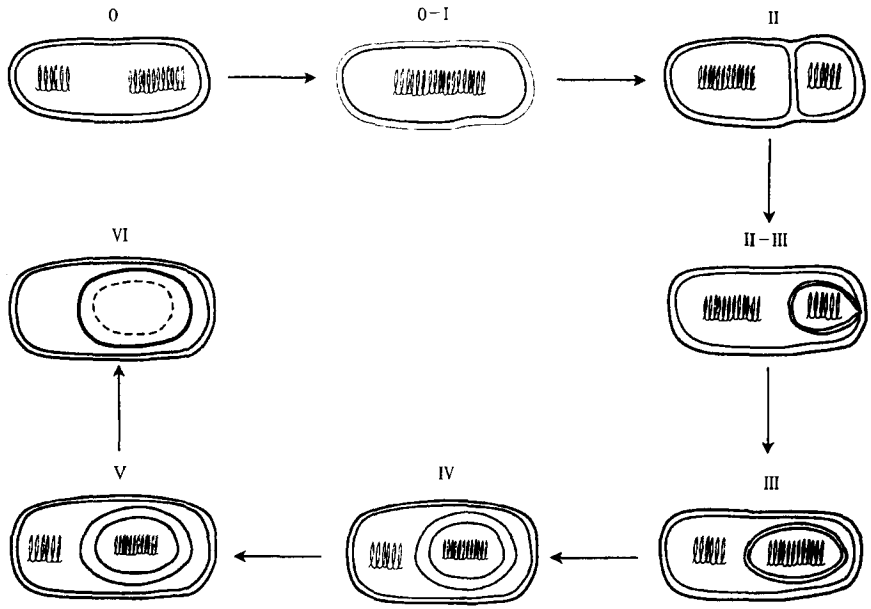


图 8－10　枯草杆菌芽孢形成阶段

芽孢形成的整个分化发育过程是由一系列基因介导的，这些基因有序而又协调地表达，并由基因产物协同完成分化发育过程。研究比较深入的有枯草杆菌芽孢形成的过程。

枯草杆菌在不利条件下会产生芽孢，芽孢的形成十分复杂，至少有 6 个因子和约 100 个基因参与。σ 因子有序地取代并在两种细胞型中协调和交叉地调控表达，它在分化过程中起主导作用。芽孢形成从不对称的隔膜发育为前孢子区和母细胞开始。σA 和 5 个 σ 次级因子依次表达、活化和取代介导约 100 个基因的表达（图8－11），这些基因按顺序开启或关闭，一些基因在前体孢子中表达，另一些在母细胞中表达。芽孢形成中最早活化的 σ 因子是 σH，它由 SPOH（或 sigH）基因编码。σH 存在于营养体细胞中，在芽孢形成启动后其浓度在很短时间内可上升几倍，这是由饥饿信号转换为基因调控信号在转录水

平上正调控表达。σH 可启动 SPOVG 及随后的 SPOⅡA 类基因的表达。SPOⅡAC（或 sigF）基因编码的产物 σF 因子是芽孢形成中继 σH 后起作用的因子。SPOⅡAC 是三顺反子芽孢形成操纵元 SPOⅡA 的三个成员（SPOⅡAA、SPOⅡAB、SPOⅡAC）中离启动子最远的一个成员。SPOⅡAC 突变后，形成异常分隔的孢子囊，说明 σF 在不对称隔膜形成方面是必要的。在 σF 的指导下，RNA 聚合酶转录第一批芽孢形成基因而不再转录营养体基因。由于 σF 的量较少，因而它只能改变部分 RNA 聚合酶的转录性质，营养体状态下的基因仍有部分被转录。σF 活化后，会导致两种事件发生：一是在前体芽孢中一个早期芽孢形成基因（SPOⅢG）表达，合成 σG 因子，σG 是前体芽孢中指导 RNA 聚合酶转录晚期芽孢形成基因的因子；另一事件是通过隔片（septum）中的一种蛋白质介导 σE 原活化为 σE 因子，σE 原由 SPOⅡGB 编码。接着 σE 因子启动产生 σK 因子的基因转录（实际上 σK 的产生很复杂，因为其基因必需通过重组才能产生），σK 亦是以无活性的 σK 原的形式存在，可被一种蛋白酶激活。σK 一旦被激活，它将取代 σE 并引起母细胞中晚期基因的转录。在前体芽孢和母体细胞中发生这些事件的计时被其他信号所控制。在分化过程中编码 σG、σK 的 SPOⅢG 和 sigK 的转录都是自动调控的，可在很短的时间内积累大量 σG 和 σK 原，以取代原来的 σ 因子，使基因表达与形态发育同步，并使母细胞区基因表达与前体芽孢区基因表达协调。

芽孢细胞形成过程中 RNA 聚合酶以新的构形行使功能，其核心酶与营养体细胞相同，仅营养体中的 σ43 因子为其他芽孢形成特异的 σ 因子取代（图 8－11），进而启动一系列芽孢形成特异基因的转录。芽孢形成整个过程大约需

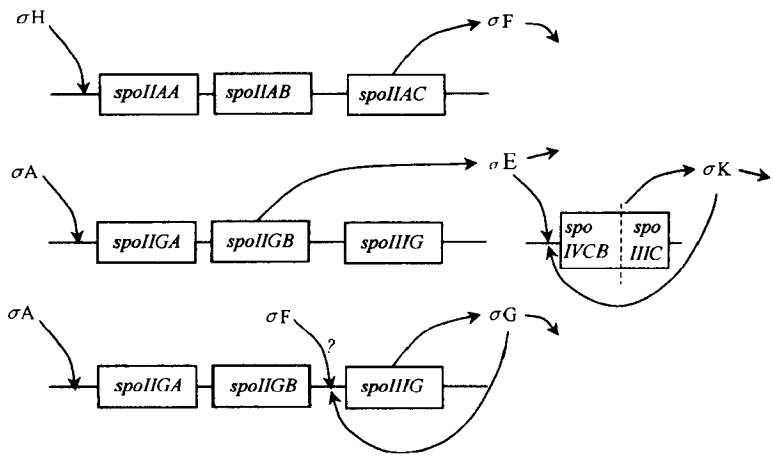


图 8－11　芽孢形成过程中的 σ 因子方框中为基因，箭头代表 σ 因子的作用或产生

8 小时左右，在这种分化中，母细胞产生两个不同的子细胞，一个最终裂解消失，另一个形成芽孢，其结构与起始细胞也不同。基因控制产孢的过程主要在转录水平，这些在营养阶段表达的基因在芽孢形成时关闭，而大多数基因继续表达，芽孢形成的特异基因只在这个时期表达，在芽孢形成的后期，40％的细菌 mRNA 芽孢形成特异的 mRNA。

二、鼠沙门氏杆菌表面抗原的转换

微生物对外界环境的适应能力是它们赖以生存的重要条件，病原微生物借助包括“模拟”宿主抗原、在吞噬细胞内存活和主要表面抗原的变异等各种机理以对抗宿主的免疫应答。广义地说，微生物抗原的变异指的是其表面成分抗原性改变，且改变率高于突变率。抗原变异的基本形式是通过 DNA 重组来调控基因表达的快速启闭而引起的双向转换。某些细菌中，两个基因交替表达产生两种类型的抗原成分，最好的例子是鼠伤寒沙门氏杆菌鞭毛蛋白 H1 与 H2 基因的交替表达（也称鞭毛相的转换）。这种转换是 DNA 水平上的控制。

沙门氏菌的鞭毛蛋白由两个不同的基因 H1 和 H2 所编码。一个细菌中若是由 H1 基因表达产生 H1 型鞭毛蛋白，这时细菌处于Ⅰ相（phase Ⅰ）；若是由 H2 基因表达产生 H2 型鞭毛蛋白，则细菌处于Ⅱ相（phase Ⅱ）。处于Ⅰ相或Ⅱ相细菌在进行细胞分裂时，以 1/1000 的频率产生另一相的后代，这个过程就叫做相变（phase variation）。相变的意义在于逃脱寄主抗体的进攻。

在一定时间内，一个细胞中 h1 基因和 h2 基因中只能有一个表达。h2 基因与另一个编码 h1 基因阻遏蛋白的基因 rh1 紧密连锁，这两个基因协调表达。当 h2 基因表达时，rh1 也同时表达，rh1 基因产物就阻碍了 h1 基因的表达。当 h2 基因不表达时，rh1 也不表达，这时 h1 基因就表达。因此，沙门氏杆菌的相变决定于 h2－rh1 转录单元是否表达（图 8－12）。

h2－rh1 转录单元是受到其上游一段 DNA 序列的排列方向所控制。这段 DNA 序列为 995bp，两端各含有一段 14bp 长的颠倒重复序列（IRL 和 IRR）。h2 的起点密码子与这段 995bp 的 DNA 相隔 16bp，但启动子位于 995bp 序列的末端。在这段序列中，一个叫做 hin 基因的编码产物可以催化这段 995bpDNA 的倒位。当启动子序列朝向 h2 基因时，则 h2－rh1 转录单元表达，而基因 h1 不能表达，细菌处于Ⅱ相。当 995bp 序列倒位后，启动子移动到另一头，而且背离 h2 基因，因此 h2－rh1 转录单元不能表达，这时 h1 基因表达，细菌就处于Ⅰ相。

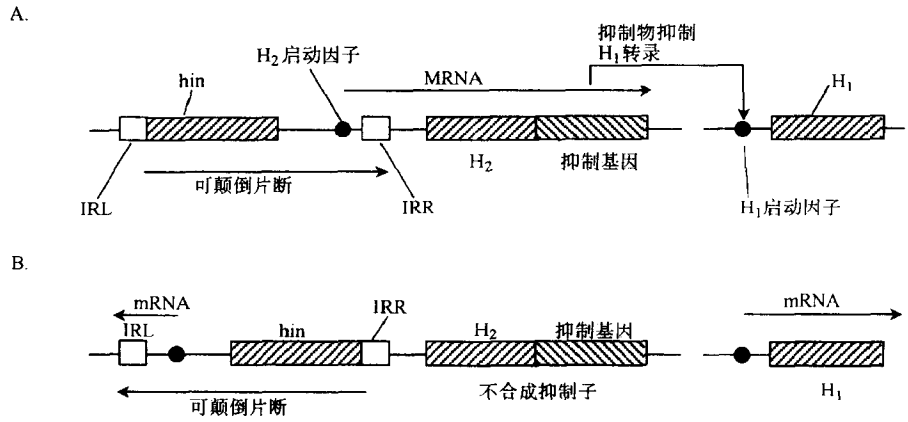


图 8－12　可颠倒片段控制沙门氏杆菌 h1 和 h2 基因交替表达

第四节　酵母交配型的转换

一、酵母的生活周期

酵母是单细胞的真菌，有单倍体和二倍体两种存在形态。具有不同交配型的单倍体细胞一方面可以通过有丝分裂进行出芽无性繁殖，另一方面又能通过两个不同交配型细胞的细胞融合进行有性繁殖。在各种酵母中，最为熟知的是啤酒酵母（*Saccharomycis cereviae*），其细胞为球状或卵圆形，直径约为 3μm。在最佳培养条件下，啤酒酵母细胞大约每 90 分钟就进行一次细胞分裂。在有丝分裂期间，每一个细胞都形成一个芽孢状附器（appendage），附器含有一整套染色体（图 8－13）。芽孢然后分离，形成一个较小的子细胞。当子细胞长大到其亲本细胞大小后，再形成芽孢。各种酵母都能以单倍体和二倍体状态存在。啤酒酵母的单倍体细胞有两种不同交配型（mating type）即两种不同性别的细胞，分别叫做 a 和 α。当 a 和 α 细胞融合后，形成 a/α 二倍体细胞。如果 a 和 α 细胞分别来自不同品系，这种接合类型叫做异宗配合（heterothalism），即两种不同类型的单倍体细胞接合，进行有性生殖的方式。酵母同一品系的 a 和 α 细胞也可以接合，产生有性孢子，这种现象叫做同宗配合（homothalism）。在适宜的营养条件下，二倍体细胞可通过有丝分裂，继续保持二倍体状态。如果营养缺乏，二倍体细胞通过减数分裂形成 4 个单倍体子细胞，叫做孢子。这 4 个孢子被包裹在一个厚的、具细胞壁的、叫子囊（ascus）的囊状体内。子囊破裂后，子囊孢子又被释放出来，最后出芽，进行新一轮单倍体世代循环。

另一研究较清楚的酵母为裂殖酵母（*Schigosaccgaromyce pombe*），其两种不同性别的单倍体细胞分别叫做 h^+ 和 h^-。这种特殊酵母并不产生芽孢，当长大到一定大小后，通过细胞分裂产生两个大小相同的子细胞。在生活周期的其他方面，裂殖酵母均与啤酒酵母相似。

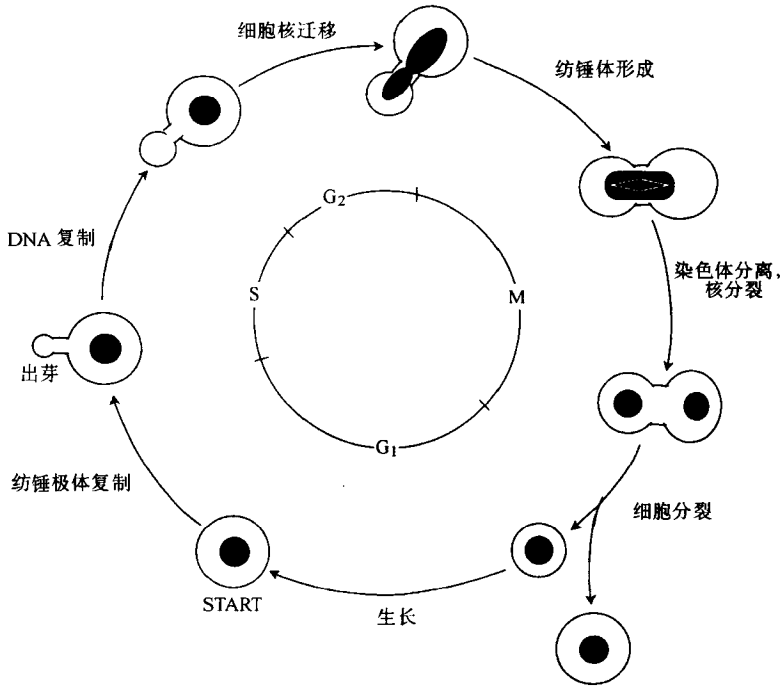


图 8－13　酵母纺锤体形成与细胞分裂

二、酵母的快速性转变

酵母的性别是不稳定的，也就是说 a 和 α 型单倍体细胞并不是绝对的和永久性的，其 a 和 α 性别可以交替变换。这种转换只发生在同宗配合的酵母品系中，而异宗配合的酵母品系则不能进行交配型转换。根据对啤酒酵母交配型基因表达的研究，在染色体 3 上有一个交配型等位基因 mat（mat gene）确定单倍体细胞属于 a 和 α 交配型。其中 mat a 等位基因确定 a 交配型，mat α 确定 α 交配型，但这两个等位基因具有不同表达方式。在同宗配合的酵母中，单倍体细胞的交配型基因能够极其快速地相互转换，所以基因的行为不符合经典遗传学的规则。由于异宗配合的品系不能进行交配型转换，所以 mat 座位的行为遵循经典遗传学的规律。现在知道，在啤酒酵母的第 4 染色体上有一个确定同宗

配合或异宗结合的 ho 座位。自然界的大多数啤酒酵母品系都是同宗配合型，异宗结合的品系通常只是一些实验室收集的突变体。

过去曾认为快速性转换可能是 mat 突然发生反转的结果，同沙门氏杆菌中控制鞭毛形成的两种蛋白质基因交替表达的方式一样。但是杂交实验表明，在染色体 3 上有一个具活性的 mat 基因和两个不表达的交配型座位 hm。其中一个不表达的座位 hml 位于 mat 座位的左侧约 200kb 的位置，含有一个拷贝的不活动的 α 信息。另一不活动基因 hmr 位于 mat 座位的右侧，含有不活动的 a 信息。这些基因的活性状况的转换与遗传信息从其中一个不表达基因转移到 mat 座位有关（图 8－14）。在这一过程中，最初的 mat 基因通过某种机制被取代，但是从一个不活动基因上转座一个盖合并不使该位点丢失任何遗传信息。因此，促进交配型转换所需的遗传信息被保存下来。酵母细胞的这种 a 和 α 性别交替转换的方式就是所谓的交配型转换的盖合模型。

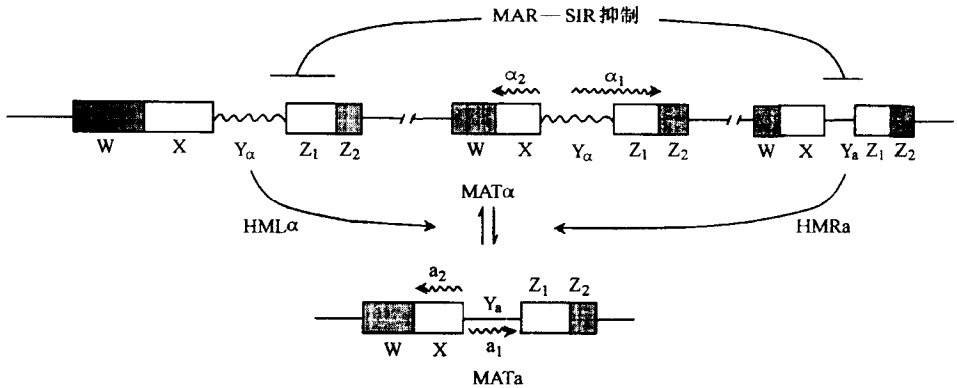


图 8－14　染色体Ⅱ上盖合的组织

HML 位于染色体Ⅲ的左臂上，HMR 位于右臂上。W 长 732bp，在 HML 和 MAT 座位之间具同源性的 X 区长 704bp，位于 HML、HMR 和 MAT 座位上。HML、HMR 和 MAT 座位只能有一个 a 特异性序列 Ya（642bp）或一个 α－特异性序列 Yα（747bp）。Z1（239bp）位于所有三个盖合之中，Z2（89bp）存在于 HML 和 MAT 之中。这三个盖合的极性相同，HML 距 MAT 200Kb，MAT 距 HMR 100Kb

在同宗配合的酵母中，a/α 细胞通过减数分裂，产生 2 个 mat a 和 2 个 mat α 孢子，但是由单个孢子长成的菌落含有 mat a、mat α 和 mat a/mat α 细胞。根据对 mat a 和 mat α 基因克隆的分析，这两个基因的结构差别很明显，在其长度为 600bp 的内部片段 Ya 和 Yα 中不存在序列同源性，但这两段序列在两端都具有共同的 X 和 Z 片段（图 8－14）。mat 座位中 Y 片段的差异是由两个不活动基因引起的，hml α 与 mat α 座位具有一个相同的 Y 片段，而 hmr a 与 mat a 座位也具有一个相同的 Y 片段。因此不活动座位具有产生两种盖合的准

确的遗传信息，也就是说，mat a 细胞能够将贮存在 hml α 中的一个拷贝的 Yα 替换 mat 座位中的 Ya 片段，而 mat α 细胞则能将贮存在 hmr a 中的 Ya 片段替换 mat 座位中的 Yα 片段。

三、酵母交配型盖合转换的调节

（一）Sir 基因关闭 hml α 和 hmr a 基因的表达

过去曾认为 hml α 和 hmr a 基因都是不活动的，因为这些基因不具有重要的启动子。后来发现 hml α 和 hmr a 都能转录出一些特殊的 mRNA，转录起点位于 Y 片段之内，而 Y 片段在 mat 座位和不活动 hm 座位中都存在同源性。因此这些不活动基因必定具有启动子。同时还发现了另外 4 个互不关联的基因 sir1、2、3 和 4 与 hm 座位表达有关，这些基因的产物一起以反式方式阻止不活动交配型基因表达。如果这 4 个基因中任何一个不表达，hml α 和 hmr a 基因的转录速率就与 mat 座位中相应基因的转录速率相同。

（二）交配型盖合运动的启始

酵母交配型盖合的转换是由 DNA 分子中的断裂缺口启始的，确切地说，它是由于在 Y 和 Z 片段之间形成双链缺口的结果。缺口是由一种位点特异性内切核酸酶切割形成的，这种酶由 ho 基因编码，叫做 Y/Z 内切核酸酶，它识别 Y 和 Z 片段之间的 CAGCTTTCCGCAACAGTAAA 序列。同宗配合品系 ho^+ 能形成这种缺口，而异宗配合品 ho^- 由于不能合成 Y/Z 内切核酸酶，所以不能进行交配型转换。交配型盖合转换的基本步骤可能都与同源染色体之间基因交换的基本步骤相同，因为酵母的交配型转换、DNA 损坏的修复以及重组过程都需要一个叫做 rad52 基因的编码产物。

双链缺口形成之后，matY 区域降解，这就是亲本 mat 等位基因丢失的原因（图 8－15）。为了替补丢失的 Y 片段，Z 片段的两个游离末端以及在除掉 Y 片段时所形成的游离末端都侵入 hm 基因内适当同源区中，然后从侵入双螺旋的断裂链的 3′端开始 DNA 修复合成，以修复丢失的 Y 片段。因此，mat 座位通过侵入 hm 座位，从 hm 座位获得遗传特异性。此时，两个交配型座位通过交叉连接在一起（图 8－15），当交叉被切割以后，两个座位分开。

Ho 内切核酸酶识别和切割的位点不仅存在于 mat 座位中，而且也存在于 hml 和 hmr 座位中，但是在活体内，这种酶只切割 mat 中的位点。sir 基因的产物与 H_O 内切核酸酶区别不同位置的切割位点有关，可能该基因的编码产物使 hml 和 hmr 区域的 DNA 浓缩，关闭这些座位的转录，同时也保护这些座位免遭内切核酸酶的切割作用。

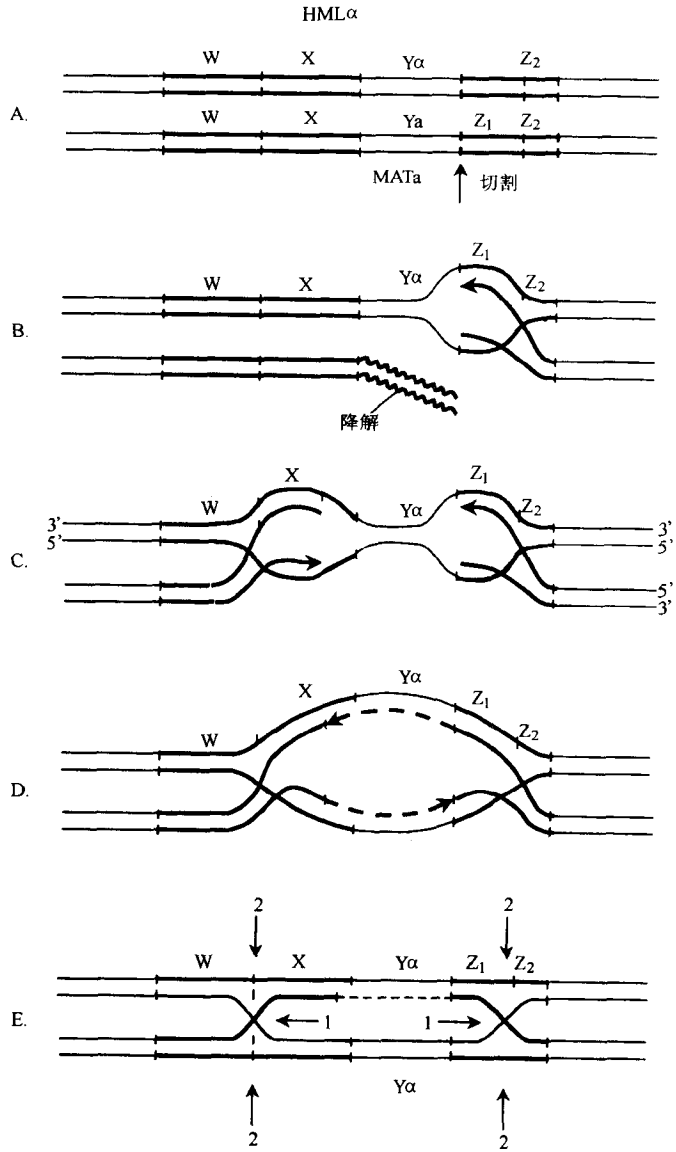


图 8－15　盖合转换和机制

A. 切割双链并降解 Y 片段　B、C.Z 和 X 片段的游离末端侵入相应 hm 基因的同源区，形成一段短的杂双螺旋区　D. 在 DNA 修复合成时，3′方向延长，以侵入的 hm 基因的 Y 片段替换丢失的 Y 片段。与盖合转换有关的 hm 和 mat 基因位于同一染色体上。箭头 1 和 2 示霍利迪结构的两种切割方式（根据 J.N.Strathern，1988）

（三）交配型盖合运动的方向

酵母交配型转换是一种高度有序的过程，只局限于单倍体亲本细胞之中，而不发生于子细胞或芽孢之中，芽孢细胞必须分裂一次后才能进行交配型盖合转换。盖合转换总是成对发生，因此盖合运动发生在 mat DNA 合成之前，交配型盖合通过复制而分配到两个子细胞之中。ho 基因编码的 Y/Z 内切核酸酶是在细胞周期的 G1 期合成的，但它只在 S 期行使功能。Y/Z 内切核酸酶的活性是不稳定的，它可能在实施其功能时受到某种因素的抑制。

交配型盖合转换的方向是有序的，它总是从 a 转移到 α 交配型，或者从 α 转移到 a 交配型。mat α 座位通常倾向与 hmr 座位结合并侵入该座位之中，而 mat a 座位则倾向与 hml 座位结合。这种对性关系由 DNA 结合蛋白通过交配型座位进行直接或间接控制。

四、mat a 和 mat α 基因编码控制蛋白

α 细胞含有几种 α 特异性蛋白，而 a 细胞则不能产生这种蛋白质。同样，a 细胞也具有一组 a－特异性蛋白，α 细胞也不能产生这种蛋白质。编码 α－特异蛋白的基因只能在 α 细胞中转录。交配型座位控制一组细胞型特异性基因即 mat a 和 mat α 基因的表达，这两组基因的编码产物活化或抑制酵母基因组中其他基因表达。因此交配型座位是细胞发生特化的一个主控制座位。交配型座位的编码产物控制某些特殊基因表达的方式如图 8－16。在 α 细胞中，mat α 座位含有 2 个基因即 mat α1 和 mat α2，其中 mat α1 编码的 α1 蛋白启动 α－特异性基因表达。相反，在 a 细胞中并不存在由 mat a 基因编码的活化蛋白，所以 a 特异性基因进行组成型表达。因此，a 特异性基因的表达只有在 α 细胞中才被关闭。关闭 α－特异性基因表达式的是 mat α 座位中第二个基因编码的 α2 蛋白。α2 为一种系列位点特异性 DNA 结合蛋白，识别每一个 a－特异性基因上游的目标位点，即 a－特异性基因的启动子，所以它是 a－特异性基因表达的负调节子。

在二倍体细胞中，α－特异性基因必须关闭，同时也必须开通减数分裂和芽孢形成途径。这两个过程由 α2－a1 蛋白质复合体调节。这种复合体蛋白控制 a/α 二倍体细胞的许多特性，并关闭一系列有关基因的表达，如通过抑制编码 α1 蛋白的基因的转录而关闭其他几个 α－特异性基因的表达。所以 a/α 二倍体细胞不表达 α－特异性基因。a1－α2 复合体蛋白也关闭 ho 基因转录，所以 a/α 细胞不能进行交配型盖合的相互转换。另一方面，a1－α2 蛋白还通过抑制另一基因 rme1 的转录而启动减数分裂和出芽过程。RME1 蛋白在 a 细胞和 α

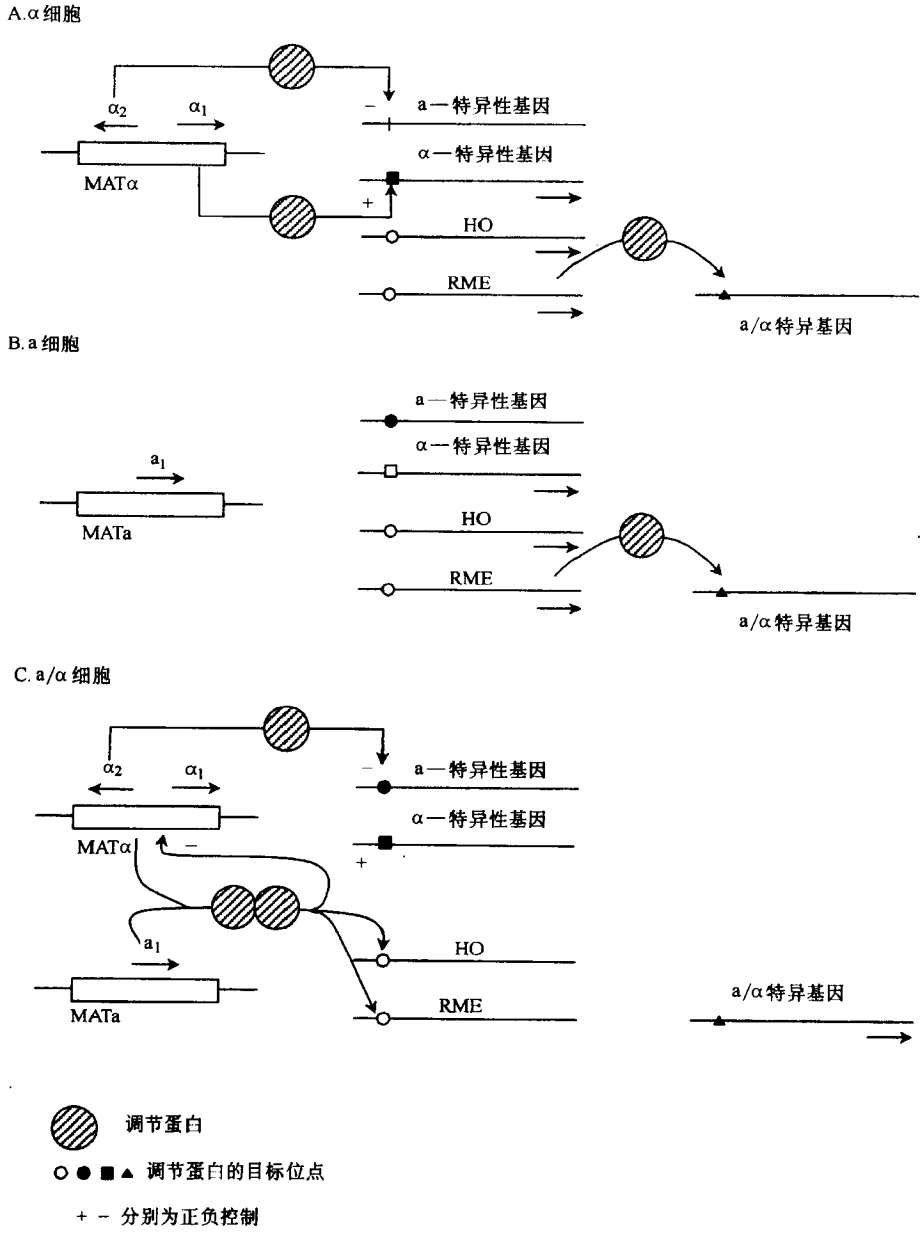


图 8－16　α单倍体、a单倍体和a/α二倍体中mat基因的功能

细胞中都能合成，为一种减数分裂的抑制子。a/α细胞不能合成RME1蛋白，所以二倍体细胞就能进行减数分裂和形成芽孢。a1－α2复合体还调节其他一

些基因的表达，但其调节机制目前都还不清楚。

总的说来，由酵母交配型座位编码的三种调节因子 α1、α2 和 a1－α2 复合体为研究不受环境条件影响的调节蛋白的作用机理提供了重要线索，是否所有高等真核细胞都同酵母一样，都具有主调节座位和主调节蛋白，并以此来维持细胞的分化，这一问题尚有待深入研究。

五、外激素与单倍体细胞的交配感受性

如果将 a 和 α 细胞混合，二者并不能自动结合在一起和融合成 a/α 二倍体细胞。这两种细胞在接合之前必须首先相互转换成对交配具有感受性的细胞，然后才能结合。α 和 a 细胞必须首先释放一种特殊的叫做外激素（pheromone）的分子，然后外激素再结合到相对的交配型细胞表面，此时这两类细胞才具有交配感受性。α 细胞向培养基释放长 13 个氨基酸的 α 因子，然后 α 因子与 a 细胞表面一种特殊受体结合，并诱导 a 细胞产生长 11 个氨基酸的 a 因子。a 因子又与 α 细胞表面的特殊受体结合。此后再通过一系列生化反应，促进 α－a 细胞成对地结合在一起，并使细胞周期停止在 G1 期。在这一时期，两种细胞均不能复制其 DNA，但能够合成其他一些主要的细胞内分子。最后这些反应在细胞壁和细胞膜中造成一些特殊变化，使成对的细胞发生融合。

α 外激素由 mfα 基因编码，a 外激素由 mfa 基因编码。mf 基因编码的初级蛋白并不具有活性，当通过特殊的水解蛋白酶的降解之后才形成较短的、具有生物活性的外激素肽链（见第 2 章）。显然，这些分子量较小的外激素便于通过细胞壁到达目标细胞膜的外表。同其他肽激素行使功能的机制一样，外激素与插入到目标细胞质膜中的特殊受体分子结合。

第五节　锥虫表面抗原转换的基因控制

一、锥虫生活史及表面抗原转换现象

锥虫是一种具鞭毛的原生寄生动物，由舌蝇或称采采蝇传播，侵入人体或其他哺乳动物，引起嗜睡症。在非洲和南美，锥虫常感染奶牛，严重降低牛奶和牛肉产量，而且还会增加动物感染其他疾病的危险性。因此对这种疾病的治疗受到了广泛重视。锥虫的生活史在舌蝇和哺乳动物两类寄主间交替完成。其整个生活史中，形态和生化特征方面变化很大（图 8－17）。锥虫的原生质膜上覆盖了一层包被，主要由表面糖蛋白构成，分子量为 $5\sim10\times10^6$，这种糖蛋

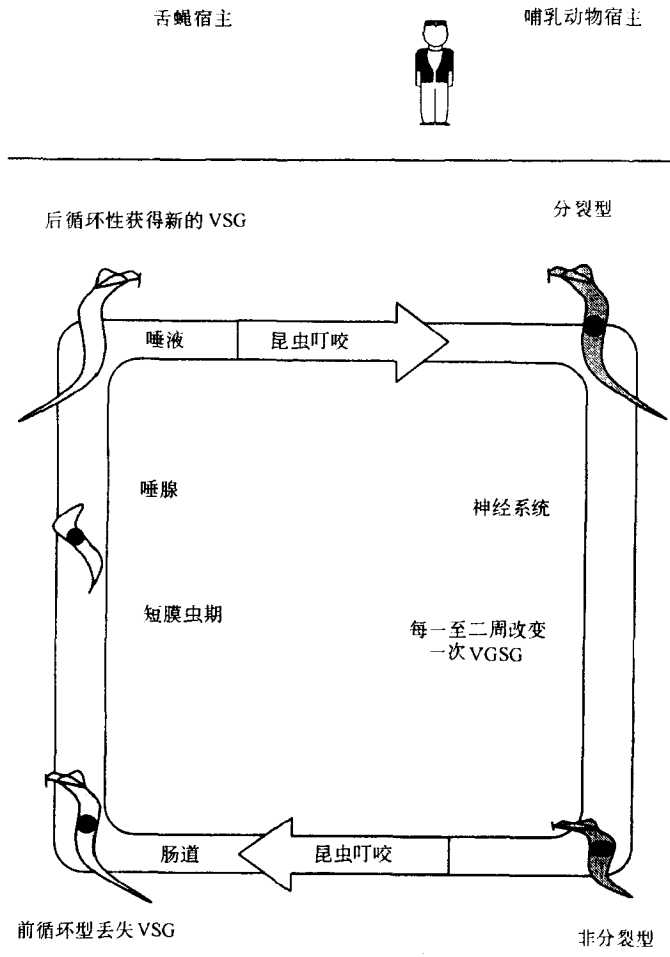


图 8－17　锥虫生活周期的形态变化

白构成了锥虫惟一的表面抗原。任何锥虫个体在特定时间里只表达一种表面糖蛋白变体（variable surface glycoprotein VSG）。寄主免疫系统能识别 VSG，将其作为外源抗原并合成其相应的抗体，将这种寄生物杀死。但是，锥虫的基因组含有数千种不同的 VSG 基因，能合成许多种不同的 VSG，因此根本就不可能制造常规的疫苗来防治锥虫感染，因为任何一种疫苗都只与一种 VSG 进行反应。锥虫在舌蝇－哺乳动物间的整个循环中通过不断地改变 VSG 以避开寄主免疫系统的识别，得以存活。当一只健康的舌蝇叮咬一个受锥虫感染的哺乳

动物时，锥虫以一种“前循环型”（procyclic form）进入舌蝇的胃部，失去VSG，大约3周后，产生的后代分化成“后循环型”（metacyclic form）锥虫，重新获得一种新的VSG，构成外包被覆盖于原生质膜。携带“后循环型”的锥虫的舌蝇叮咬哺乳动物时，将锥虫传染给哺乳动物，锥虫在哺乳动物的血液中繁殖，在侵染后的1周内继续表达“后循环型”的VSG，1～2周后则合成另一种新的VSG，以后每隔一周VSG转换一次。

一个刚合成的新生VSG分子约由500个氨基酸构成，包括N－末端的信号序列（约20个氨基酸）、抗原决定子（360个氨基酸左右），和C－末端的同源区域（约100个氨基酸），同源区域后连接一个疏水的尾部结构（约20个氨基酸）。两端经加工后形成成熟的VSG，加工过程中信号序列被切除，疏水尾部在VSG到达膜表面时被移去，暴露出新的C－末端以共价键结合于锥虫膜上，表达抗原作用。每一种VSG的免疫学特征都是不同的，哺乳动物免疫系统根据锥虫表面抗原的转换也不断作出反应，但这种反应滞后于锥虫表面抗原的变化。VSG的频繁转换使锥虫不断逃过寄主免疫系统的监查，得以存活，同时伴随着一次新的寄生物繁殖高峰，引起发热和皮疹，寄生物最后侵入到哺乳动物中枢神经系统，引起寄主昏睡，最终昏迷。

二、锥虫表面抗原转换的机理

如前所述，鼠伤寒沙门氏杆菌能颠倒一小段DNA片段而将一种表面抗原转换成另一种抗原。但是，锥虫至少可以在100种不同VSG之间进行转换。显然，简单的DNA片段颠倒并不能解释锥虫的这种转换行为。酵母交配型转换和锥虫可变表面抗原转换之间存在很大相似性。hmr和hml上的静止的交配型盖合类似不转录的基本VSG基因拷贝的衔接重复。基因从不表达位点hmr或HML转移到mat（表达位点）位点类似抗原转换。可能在其他方面还存在相似性。在酵母中交配型转换不仅需要mat座位的遗传信息进行表达，而且hmr和hml座位有选择地受到抑制是由sir基因产物控制的。这种基因产物可能通过改变静止盖合周围的染色体结构而对这两个盖合的功能进行调节。根据活性染色质和非活性染色质之间的差别，可能锥虫一次只表达一个端粒VSG基因的机制与酵母相同。

事实上，锥虫的抗原转换要复杂的多。首先，这种原生动物比细菌大的多，而且在任何时候都能表达某种特殊的VSG，而这种VSG至少占锥虫总体蛋白的10%。为了合成含量如此之高的VSG蛋白，VSG既可由多基因编码，也可以一种特殊的速率进行转录，还可以一种特殊的速率对VSG mRNA进行

翻译；第二，锥虫必须很快地转录 VSG（估计每世代约为 10^{-4}至 10^{-5}）。因此，在一个表达一种 VSG 抗原的锥虫群体中，至少有一个个体在寄主免疫系统杀死所有锥虫之前转换成另一种新的 VSG 抗原。但是，表达新 VSG 抗原的锥虫就不必继续表达老的 VSG 抗原，否则，它们就受老抗体的攻击。只有 VSG mRNA 不稳定才能实现这种快速的抗原转换，而 VSG mRNA 的不稳定性又需要相应的 VSG 基因进行快速转录，合成大量具有翻译活性的 VSG mRNA。

目前研究的最清楚的锥虫是 *Trypanosoma brucei*。将此锥虫第一次接种家兔后，由于载有特殊 VSG 的 *T. Brucei* 在血液中繁殖而使家兔致病。当家兔对第一次 VSG 作出反应后，家兔恢复健康，但不久后，由于含有新 VSG 的锥虫，家兔重新患病，最后这种新的 VSG 成为血液中主要的 VSG。在寄生生活的第一、二、三周期之后，通过克隆血液中寄生虫 mRNA，有可能获得相当于一个细胞系中连续地从一个 VSG 基因转换到另一基因的各种 VSG mRNA。当利用这些 cDNA 为探针，与三个连续世代的寄生虫群体的 DNA 进行杂交后，就可知道 VSG 基因在基因组中的组织（图 8－18）。

锥虫的基因组组成与一般有机体不同，由大量独立的染色体单元组成，具有 8～10 条长 200 至 2000Kb 的大染色体和数百条长度在 50～150Kb 的小染色体。每一种 VSG 基因都是由一个基本拷贝基因（basic copy gene）编码，VSG 基本拷贝基因是一个庞大的基因家族，按照基本拷贝基因在染色体上的位置，可以分为端粒基因（telometric gene）和内域基因（internal gene）两类。端粒基因位于染色体端粒区 5～15Kb 范围内，假如每个端粒有一个基因，则至少有 200 个端粒基因；内域基因位于染色体内，距离端粒在 500Kb 以上。各个基因间都有一定的关系，两个密切相关的基因可能激发相同的抗原反应（免疫），称之为同功基因（isogene）。虽然每一个锥虫个体都带有所有 VSG 基因组分，但在发育的特定时间，一个锥虫个体只选择其中一个 VSG 基因进行转录表达。该基因的拷贝称为连锁表达拷贝（expression－linked copy 简称为 ELC），连锁表达拷贝都位于端粒附近。

在如此庞大的基因家族里，锥虫选择连锁表达拷贝基因改变 VSG 的途径有两个：

1. *表达位点保持不变，而连锁拷贝变了*　一个基本拷贝基因序列通过复制转移到表达位点，取代正在表达的基因序列，使本身被激活，这种一个“盖合”取代另一个“盖合”的过程对基因位点本身没有影响，但表达的基因拷贝序列变了。无论是端粒基因，还是内域基因都能以这种途径转移到表达位点而被激活，从而形成新的 VSG。

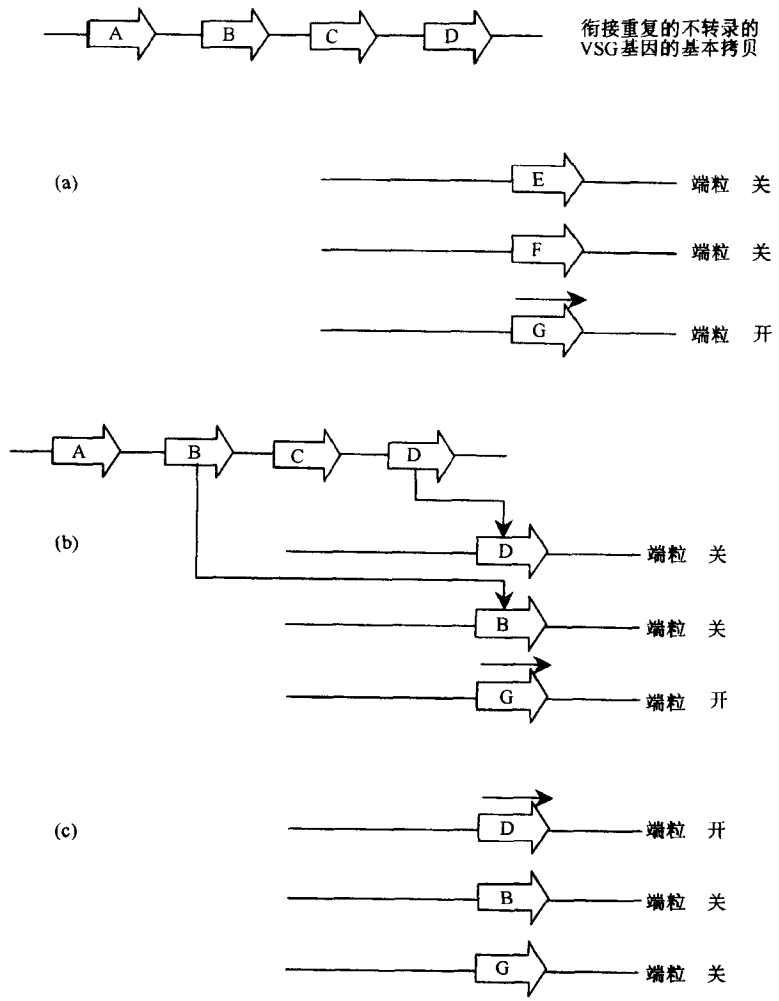


图 8－18　锥虫 VSG 基因在端粒表达位点上的活化和替换过程

A. 感染初期只有一种特殊 VSG 抗原表达，三个或更多的染色体具有端粒 VSG 基因，但是只有表达位点才进行转录　B. 感染后期，通过基因转化用新的可变基因替换一个或更多的 VSG 基因　C. 在第二次发病期间，旧端粒表达位点关闭，新位点活化，只有新的 VSG 表达

2. 改变表达位点　位于端粒的一个沉默的基本拷贝可以在原位被激活表达，而此时端粒区原先表达的位点失活，这种途径适合于端粒基因被激活。内域基因拷贝有时先转移到一个不表达的端粒区域，表现为端粒基因，然后也可

以在原位通过激活途径而表达。

激活内域基因都需要使其复制产生一个新的基因拷贝序列，而原来的基本拷贝仍在原处没有变化。端粒基因的激活既可原位被激活，也可以复制后转座到另一个染色体的 ELC 位点被激活。几乎所有的 VSG 转换都涉及一个预存的沉默拷贝取代 ELC。但也有例外，当 ELC 序列与基因组内任何沉默拷贝不匹配时，通过一系列的基因转变将不同沉默拷贝上的短 DNA 片段相互连接产生一个新序列，取代 ELC，这种过程类似于鸡的免疫球蛋白产生变异的过程，从而进一步扩大了锥虫 VSG 的变异。在 ELC 位点的 VSG 基因的结构也与一般基因不同，从基本拷贝转座至 ELC 的 DNA 长度约 2500～3500bp，但 VSG 编码的总长度仅 1500bp，多余的 DNA 主要位于编码序列上游。

无论抗原类型如何，成熟的 VSG mRNA 的前端都是一个相同的 35bp 的前导序列，然后以反式作用加上一个新合成的 5′端。但是编码这段 35nt 的小外显子的 DNA 序列并不是 VSG 基因基本拷贝的组成部分，端粒表达位点也不含这种 35bp 的小外显子序列。相反，在锥虫基因组的其他位置则存在约 200 个拷贝的 35bp 的序列，但主要以 1.35Kb 的衔接重复片段的形式存在。可见，VSG mRNA 的转录是不连续的，其 35bp 的小外显子序列的转录与 VSG mRNA 的基本序列的转录是相互独立的，二者在转录后再连接在一起，形成完整的 VSG mRNA。转录的起始和终止信号是由转座区序列决定的。事实上起始转录的控制序列可能位于 VSG 基因本身上游数千 Kb 以上，启动子定位于 VSG 序列上游 4～60Kb，更远的启动子产生的转录子含 VSG 和其他基因产物。经过切割释放出 VSG mRNA 序列，然后 35 个碱基的前导序列加到 5′端。内切酶不能切割在转座区两侧的区域，这些障碍区由重复 DNA 构成，向上游一直延伸 40Kb、下游延伸 8 Kb 左右，下游再往前的障碍区由大量的 CCCTAA 重复序列构成，一直延伸到端粒。上游往前，也有许多重复序列，其性质尚不清楚。障碍区的存在不利于 ELC 基因的研究和克隆。表达位点的激活与转录单位上游区间的变化有关。当一个 ELC 位点的启动子被另一个启动子替换时，该 ELC 被抑制和激活的能力没有改变，因此，不是启动子－反作用因子控制的结果，调节过程可能是通过改变染色质结构和启动子上游区间的结果。

侵染过程中 VSG 基因表达的顺序是不规则的，但也不是完全随机的，这可能是锥虫存活的重要特征。如果 VSG 基因按照某个预定的规律转变，则寄主可能利用一个较早的抗体模式来抑制侵染。产生 VSG 不可预测性是该系统进化为许多供体序列和多个受体位点的结果。

第六节　高等真核生物发育阶段的基因控制

高等真核生物个体结构复杂，不同的器官、组织和细胞等的结构迥然不同，其基因的表达遵循严格的时空秩序。目前这方面研究最清楚的是果蝇胚胎形成过程中基因表达的时空特异性。本节将主要以果蝇为例对高等真核生物发育阶段的基因控制进行讨论。

一、果蝇中编码体型决定因子的基因

果蝇发育开始时，沿受精卵的前后轴（anterior－posterio）和背腹轴（dorso－ventral）建立起形态发生梯度（morphogenetic gradient），不同的基因在不同的区域表达，结果使胚胎形成许多体节，每个体节分化为成虫的特定形态。前后体系调控幼虫各器官的位置信息，使受精卵的前端发育成幼虫的头，后端变为尾。背腹体系则调控组织的分化，如中胚层、神经外胚层、背腹外胚层的形成等。是什么机制导致了胚胎具有极性呢？有哪些基因参与了编码果蝇体型决定的因子呢？

现在知道，受精卵的梯度形态发生是由分子（蛋白质或 RNA）在胞浆中的差别形成的，这种不对称分布在卵子发生时就开始建立了。雌性果蝇有一对卵巢，每个卵巢由 16 根卵小管组成，每根卵小管又分为两个区，卵母细胞发生区和生长区。卵母细胞发生区位于卵小管的前端。在这个阶段，雌性生殖干细胞（stem cell）通过分裂产生一个子干细胞和一个包囊母细胞（cystblast）。包囊母细胞连续进行 4 次有丝分裂产生 16 个彼此靠细胞质桥（cytoplamic bridge）相连的细胞，各个细胞都有 2 个、3 个或 4 个细胞质桥连接点，其中只有一个包囊细胞具有微管组织中心，以后将决定并分化为卵细胞，称为前卵母细胞（preoocyte），其他 15 个细胞变成滋养细胞（nurse cell），卵母细胞与滋养细胞质相连的一端称为卵的前端（图 8－19）。滋养细胞内细胞质成分包括蛋白质和 RNA 等都进入卵母细胞，这些物质的积累导致卵体积增大。滋养细胞产生的一些 mRNA 在卵中的定位在空间上是不连续的，卵母细胞和卵泡细胞间的相互作用对胚胎梯度的形成也是很重要的。

母体内表达的基因对于卵的早期发育至关重要。用果蝇雌性不育突变体鉴定了许多此类母体基因，这些突变对果蝇本身没有影响，但不会有后代，所产生的卵不能发育成成虫。其胚胎由于缺乏角质层（cuticlar pattern）可以被识别，它们在发育过程中死亡。

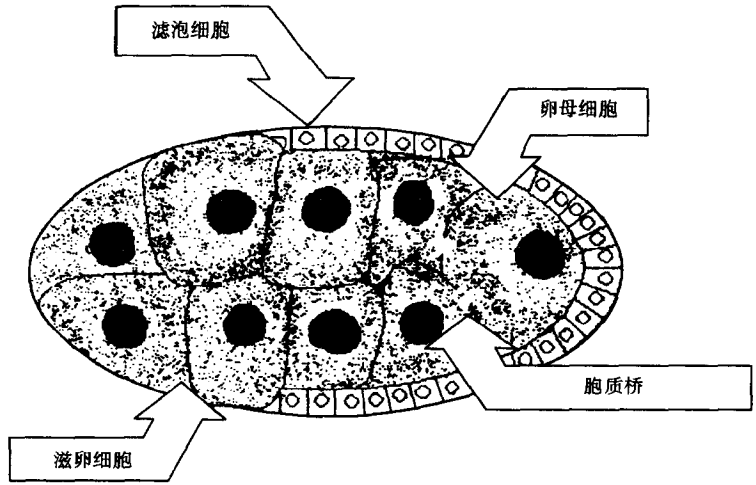


图 8－19　果蝇卵细胞

母体基因的特征是它们在受精前表达，其产物可能在表达时发生作用，也可能先贮藏起来以后再发生作用。根据母体基因的表达位置可以将其分为两类：母体体细胞基因（maternal somatic gene）和母体生殖细胞基因（maternal germline gene），前者在母体体细胞内表达，影响卵的发育，如在卵泡细胞内作用的基因；后者在生殖细胞内表达，这些基因可能在滋养细胞或卵母细胞内表达。

用母体基因突变法已鉴定了果蝇四群母体基因与胚胎特定区域发育有关。用常规遗传学试验法（双突变特性与单一突变特性的比较）和生化测定法发现每一群基因都可以组成一个发育途径，其中的基因按一定的次序进行活动。四个途径中的各个成员（图 8－20），它们都有一个共同的操作原则。每一个途径都被卵内或卵外存在的局部事件所启动，导致卵内信号的区域化。此信号是蛋白质，在卵内呈不对称分布，称为形态发生素（morphogen）。形态发生素的定义为：一种由局部浓度导致周围形成特殊结构的蛋白质。形态发生素连续的浓度梯度会引发一系列不同的细胞分化反应，即在某浓度临界值的之上可能会导致一种方向的分化，低于此临界值，则会发生向另一种方向的分化。假设胚胎中某个部位某种形态发生素的浓度导致此部位的细胞产生结构 A，而在低于此浓度的胚胎的某个部位则产生结构 B，在更远的位置此形态发生素的浓度达到另外一个临界值，则此处细胞的命运是形成 C（A、B、C 为不同的组织或器官）。形态发生素有的是转录调节因子，有的导致转录调节因子区域化激活。

四个发育途径中有三个系统可能和前后极性有关，一个和背腹极性有关。

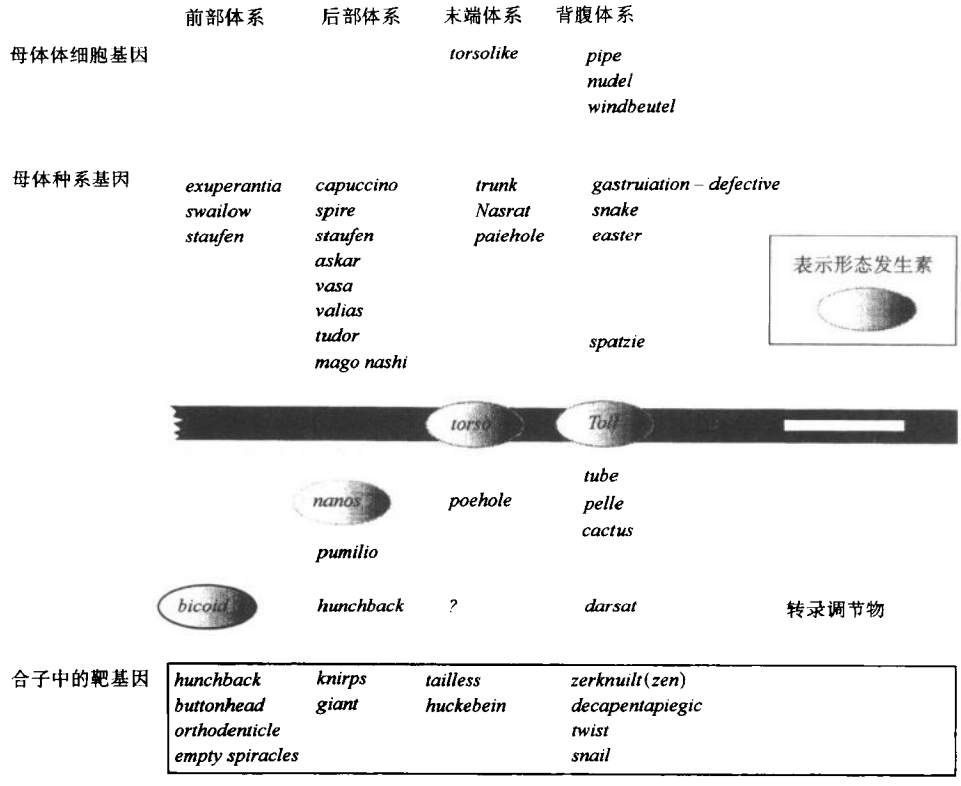


图 8－20　与果蝇胚胎特定区域发育相关的母体基因

1. 前极系统（anterior system）　前极系统负责头胸部的发育，此体系的形态发生素为 Bicoid 蛋白。Bicoid 蛋白是第一个从分子水平上被鉴定的形态发生素。Bicoid 基因的 mRNA 是在母体滋养细胞内合成的，通过细胞质桥转运到卵母细胞中，定位于胚胎的前部区域（前极），其编码的蛋白为一种 DNA 结合蛋白，其基本结构单元在其他多种生物的对发育起重要作用的蛋白中均可找到。Bicoid 蛋白一旦在胚胎前部末端合成，就会在胚胎前后轴方向上形成梯度。利用拯救（rescue）技术已证实了 Bicoid 蛋白决定前极系统的发育，如母体是 Bicoid 基因突变体，则产下的卵发育成无头或无胸的幼虫。抽取野生型胚胎前极区的细胞质注射到突变体卵的前端可以使幼虫发育恢复正常，恢复的程度取决于野生型胚胎的细胞质的量，这种表达方式称为异位表达（ectopic expression）。作为细胞质供体的野生型胚胎，其细胞质作用的效率取决于供体本身携带的 biciod 基因的数目，这表明野生型胚胎的前极区含有某些物质的浓度

取决于 bicoid 的基因剂量。实验室条件下，只有提纯的 bicoid mRNA 可以替代前极区的细胞质。

Bicoid 蛋白是一个转录调节因子，其浓度决定着胚胎前部细胞的发育结局。它特异地启动一些基因的转录，其中一个主要的基因就是 hunchback。hunchback 基因是一个合子型基因，其转录依赖于 Bicoid 蛋白的浓度，即 Bicoid 蛋白浓度高于一个特定阈值时，胚胎中该区域的 hunchback 基因才开始转录。Hunchback 蛋白在胚胎中的梯度分布与 Bicoid 蛋白是平行的，其突变对前部结构的影响与 bicoid 突变导致的后果相同。对 hunnback 转录起始点上游启动子的分析发现其包含 Bicoid 蛋白的三个高亲合力区域和三个低亲合力区域，进一步通过对带有受高或低亲合力 Bicoid 蛋白结合位点启动子驱动的报告基因的转基因果蝇研究表明，位点的亲合力决定了 Bicoid 蛋白使基因转录激活的阈值浓度。

2. *后极系统*（posterior systerm）　后极系统负责腹部分节，大量生殖细胞产物的共同作用导致 nanos 产物定位于卵的后端。nanos 蛋白是一种后极系统形态发生素，通过 mRNA 翻译控制，局部抑制 hunchback 的表达。后极系统的形态发生素作用与前极基本相似，母体基因 Nanos 表达的 mRNA 定位于后极，与 Pamilio 蛋白联合起作用。后部系统中局部定位更复杂，因为源于滋养细胞的后极决定因子必须穿过整个卵原细胞到后极。Nanos 蛋白的作用是抑制母体 hunchback 转录因子的翻译，进一步抑制那些干扰后极发育的基因表达。

bicoid 和 nanos 基因都对 hunchback 基因的表达有作用。hunchback 编码一种转录抑制子，它的存在是头胸结构形成所必须的，它本身有一种复合表达模式。在卵子形成过程中，它开始转录形成的 mRNA 均匀地分布于卵子内。受精后，Hunchback 模型按两种方式改变。Bicoid 梯度激活 hunchback RNA 在前极区的合成，Nanos 梯度阻止 hunchback mRNA 在后极区的翻译，这种抑制导致了 mRNA 的降解，前后极两个系统的协调同时提高了 Hunchback 在卵的前半部分的水平，减少后半部分的含量。hunchback 基因与其他几个与腹部形成相关的基因统称为体节缺口基因（gap gene）。这些基因在受精后两小时内表达，hunchback 基因最先表达，产生的蛋白是一种形态发生素，是其他几个基因的转录激活或抑制因子，这些基因包括 krüpple，knirps 和 giant，它们也编码转录因子，调控胚胎中的后续发育事件。Hunchback 蛋白控制 krüpple 基因表达的前后界线，高浓度时，抑制 krüpple 的表达，而 Hunchback 蛋白浓度低于一个阈值时，则又可作为转录激活因子激活 krüpple 的表达。knirps 和 giant 在前极的表达受 Bicoid 和 Hunchback 蛋白的调节，即 Bicoid 可激活其表达，Hunchback 则抑制其表达。此两基因在后极边缘的表达则依赖于另外一个体节

缺口基因 tailless。

3. 末端系统（terminal systerm）　这一系统负责胚胎两端不分节部分的一些特殊结构的发育。如头部的顶节 和尾部的尾节（telson）。这一系统的作用机制尚不清楚。现已鉴定出几个母体基因参与胚胎前后末端结构的发育调节，Torso 基因便是其中一个，此基因编码一个越膜受体蛋白，具有酪氨酸激酶活性。尽管 Torso 蛋白均匀分布于囊胚细胞表面，但其作用的位置仅限于胚胎前后极区域。Torso 蛋白受体的定位激活会引发一系列细胞内信号级联反应并最终导致各种合子型基因的表达，其中包括 gap 基因如 tailless，此基因编码一种类似转录因子的类固醇受体。这种由细胞表面信号引发的信号传导的级联反应在进化中具有很高的保守性。

4. 背腹极性系统（dorsal－ventral systerm）　该系统主要负责背部和腹部的发育，它由腹极的滤泡细胞发出的信号所启动。不同组织在背腹方向上的分布称为果蝇胚胎的背腹图式。背腹图式形成的途径是由背腹极性系统基因在转录和翻译水平上进行调控，以及通过细胞间相互作用等共同组成的。在这些基因的作用下，背腹轴上不同水平的囊胚细胞分别分化为不同的组织。

胚胎背腹极性的形成需要至少 11 个母体基因的编码产物，如果其中任何一个基因突变都会导致胚胎腹部有背部结构。通过注射野生型胚胎的细胞质，可使背部化基因的某些突变体得到部分或完全恢复。使突变体得到恢复的物质可能是 mRNA。由于来自一种背部化基因突变体的胚胎的 mRNA 可以使另一种背部化基因发生突变的缺陷胚胎得到恢复，因此，这 11 个基因都是形成背腹极性所必需的。

胚胎背腹水平上不同组织的类型是由背腹母体效应基因通过影响单个形态发生素而决定的。遗传学和实验胚胎学方面的证据提示，dorsal 基因是决定胚胎背腹图式的关键基因。在卵子发生过程中，dorsal 基因在滋养细胞中转录和翻译，产物转移到卵内。从合胞体囊胚到细胞囊胚形成过程中，Dorsal 蛋白呈不对称分布，在腹方最高，背方几乎没有。高浓度的 Dorsal 蛋白区域产生腹方中胚层，腹方两侧的中浓度的 Dorsal 蛋白决定神经外胚层，更低浓度的 dorsal 蛋白产生背方两侧结构，无 dorsal 蛋白导致背方外胚层发育。

背腹极性系统中的其他基因在背腹图式形成过程中也起着重要作用。如滤泡细胞中的 3 个母体效应基因 pipe、nudel 和 windbutel 参与了腹方信息从体细胞传递到生殖细胞的过程，它们的功能是传递一个信号给卵细胞。此信号的本质尚不清楚，但已知它引起另外几个基因（snake，easter 和 spatzle）的相互作用，snake 的产物使 easter 产物裂解而活化 spazle，活化的 spazle 产物则是 toll 基因编码的受体的配体。Toll 蛋白不对称活化导致 Dorsal 蛋白梯度分布。toll

基因的作用将在下面讨论。

二、果蝇的 toll 基因与形态发生梯度

果蝇背腹图式形成过程中，toll 基因（toll gene）编码的蛋白是该途径在卵母细胞中起作用的第一个成分。Toll 蛋白在腹侧有活性而在背部没有，从而在胚胎中形成背腹梯度。toll 基因的隐性突变体产生缺少腹侧结构的胚胎，其显性突变体则使腹面结构扩展到胚胎的背面，这说明 toll 基因的产物可能是一种形态发生素，或是一种与形态发生素有关的因子，这种因子使卵形成腹面结构。显性 toll 突变体会导致在正常的腹面位点过量合成 toll 基因产物，或者使整个卵粒异常表达 toll 基因产物，所以就使细胞在错误的位置形成腹面结构。

Toll 蛋白发生作用的机理尚不十分清楚，但已知它是通过 tube 和 pelle 两基因而起作用。tube 的功能也不清楚，pelle 编码一种激酶，其作用的靶物质可能是 cactus 的产物，此产物调控着 dorsal 编码的转录因子。在胞质中，Catus 蛋白和 Dorsal 蛋白形成复合体，使 Dorsal 蛋白不能进入细胞核中。当 Catus 蛋白被 Pelle 蛋白磷酸化后，它将 Dorsal 蛋白释放出，Dorsal 蛋白即可进入核中。可见，Toll 活化的结果是在胚胎中由腹侧至背侧建立起 Dorsal 蛋白在核中活性的梯度。在腹部，Dorsal 蛋白进入核，在背部则停留在细胞质中，使腹部结构发育所必需的基因 twist 和 snail 被活化，而背部发育的基因 dpp 和 zen 在腹部受到阻遏，仅在背部表达，发育出背部结构。

综上可见，形成果蝇 4 个发育途径的基因成员都是母体的，所以建立起始模式形成的 4 个系统在受精前已经发生，体躯两个轴的形成是独立的。影响极性的突变可能引起尾部区域发育成头部或腹部区域发育成胸部。在前后极性系统中，前极和后极中控制胚胎阶段发育的物质呈两个相反的梯度，任一梯度的缺乏都影响体节发育，而末端系统和背腹系统的作用是各自独立的。

在编码体形决定因子的基因表达后，合子的分节基因开始表达，对体轴和分节发育起作用，从而奠定了下一步发育的格局。

三、果蝇器官分化的同源异型基因

前已述及，果蝇胚胎发育中最早起作用的基因是母体基因，它们决定了胚胎的前后和背腹极性，从而使位于胚胎内不同位置细胞核的基因被选择性地激活。这些基因是处于发育第二层次的基因，称为分节基因，它们对体轴和发育进一步起作用，从而奠定分节的大格局。各体节的形态特征（头、胸或腹）的

进一步发育则依赖于第三层次的基因－同源异型基因。

早在1894年，Bateson在果蝇的杂交互补实验中观察到一类同源异形(homoesis)现象，即果蝇的某个基因发生突变会导致肢体的某一部位的器官错误地、特定地长在其他部位器官应该长的位置上。例如，触角antp（antennapedia，触角）基因突变导致果蝇第二胸体节的腿长在头部应该长触角的部位，bx（bitorax，双胸）基因突变导致果蝇后胸发育成前胸，由bx、pbx（postbitorax，后双胸）同时发生突变，会导致胸体节重叠而长成4个翅膀、8条腿。上述现象常常被称为同源转化（homeotic transformation），将这种能导致同源异形现象的基因称为同源异型基因（homeotic gene）。

果蝇的同源异型基因在染色体上常成簇地沿直线排列，称为复合座位(complex loci)，在果蝇成蝇发育过程中至少涉及到两个这样的复合座位，即触角族复合座位（antennapedia complex，ANT－C）调节头和前部的胸节发育；双胸复合座位（bithorax complex，BX－C）调节后部的两个胸节和腹节的发育。它们都含有几个同源异型基因，这些基因在染色体上的遗传顺序与其在成蝇体内表达的位置有关。

果蝇最前端的4个准体节（parasegments 1～4）的特征是由复合型基因座位ANT－C所决定的，该复合基因座包括有lab（labial，唇）、pb（proboscipendia，喙）、dfd（deformed，小眼畸形）、scr（sex combs reduced，性梳减少）、ftz（fushi taraazu，少体节）、antp等基因，它们成串排列于一个约350kb的区域中，另几个在不同发育阶段表达的基因如zen（zerknullt，视叶）、bcd、ama（amalgarm），相间于该区域（图8－21），最左侧的lib基因作用于最前端，影响头部，最右侧的antp基因作用于后部的T2－T3体节。antp基因约103kb长，有8个外显子和很大的内含子，实际编码蛋白质的DNA部分仅占1%。这一基因有两个启动子，一个位于外显子1的上游，另一个位于外显子3的上游，两者相距约70kb。转录的终止位点也有两个，但无论转录从何处开始和终止，编码的蛋白质是相同的。antp蛋白通过在胸部作用，促进T2－T3体节的形成，抑制头部结构形成，它的突变也可以使触角发育成第二对腿，或第二、第三对腿发育成第一对足，从而使T2－T3体节类似于前端的第一胸节T1。头部的antp过分表达，则前部体节发育成胸部结构。

BX－C复合座位包括ubx（ultrabithorax，超双胸）、abdA（abdominal A，腹节A）和abdB（abdominal B，腹节B）三个互补群，每个互补群上包含一个以上的基因。BX－C复合座与ANT－C复合座都位于3号染色体上，两者相隔10cM左右。BX－C总长约315kb，仅有1%～4%的DNA序列编码蛋白质。突变分析表明，非编码区的序列在调节不同体节发育的特异性方面起重要

作用。从 T2 至 A8 体节的发育都受 BX－C 基因的影响，因此，BX－C 复合座与果蝇身体的大部分发育有关，BX－C 复合座缺失时，昆虫在胚胎发育后期死亡。与 ANT－C 一样，这个复合位点的主要特点也是遗传图上左侧的基因影响前端，而右侧的基因影响后端（图 8－22）。

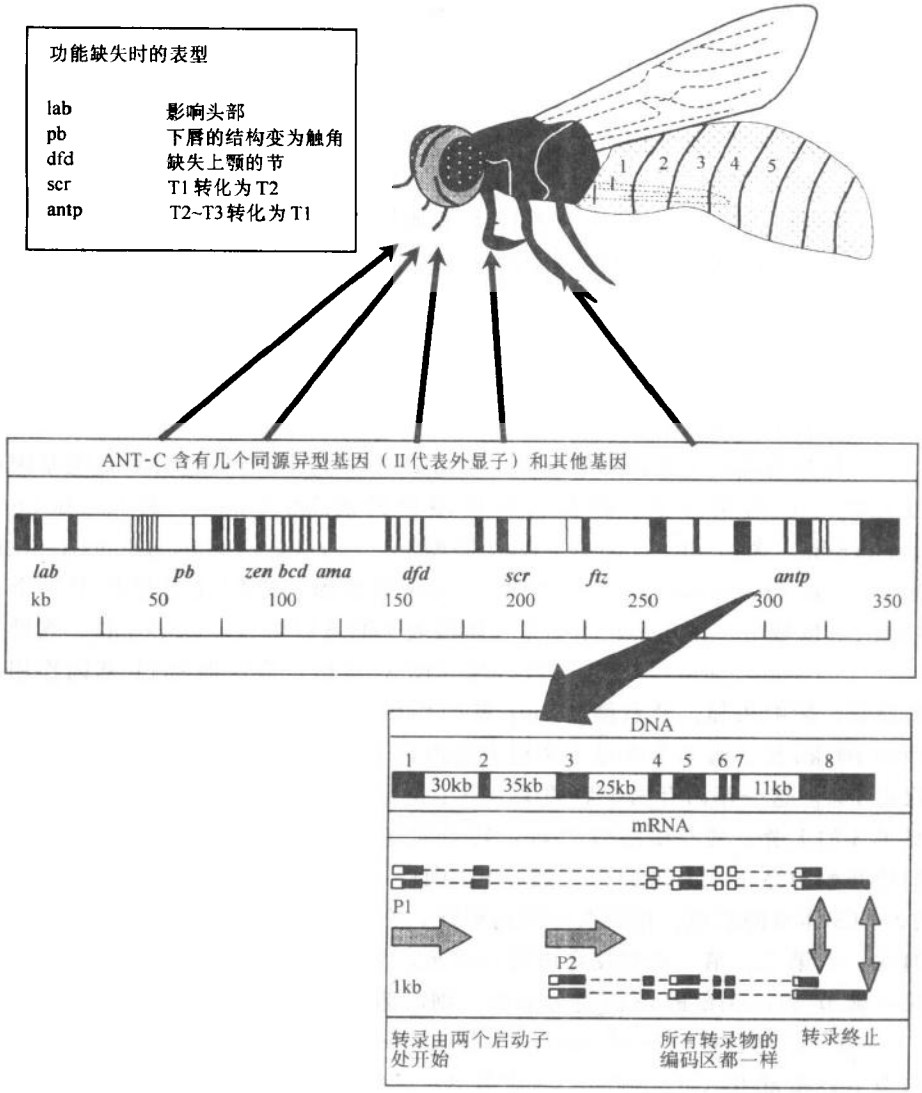


图 8－21

NAT－C 复合座位的同源异型基因决定果蝇前部各体节的发育，在同源异型基因之间还散布着其他基因。antp 基因很大，有不同表达方式

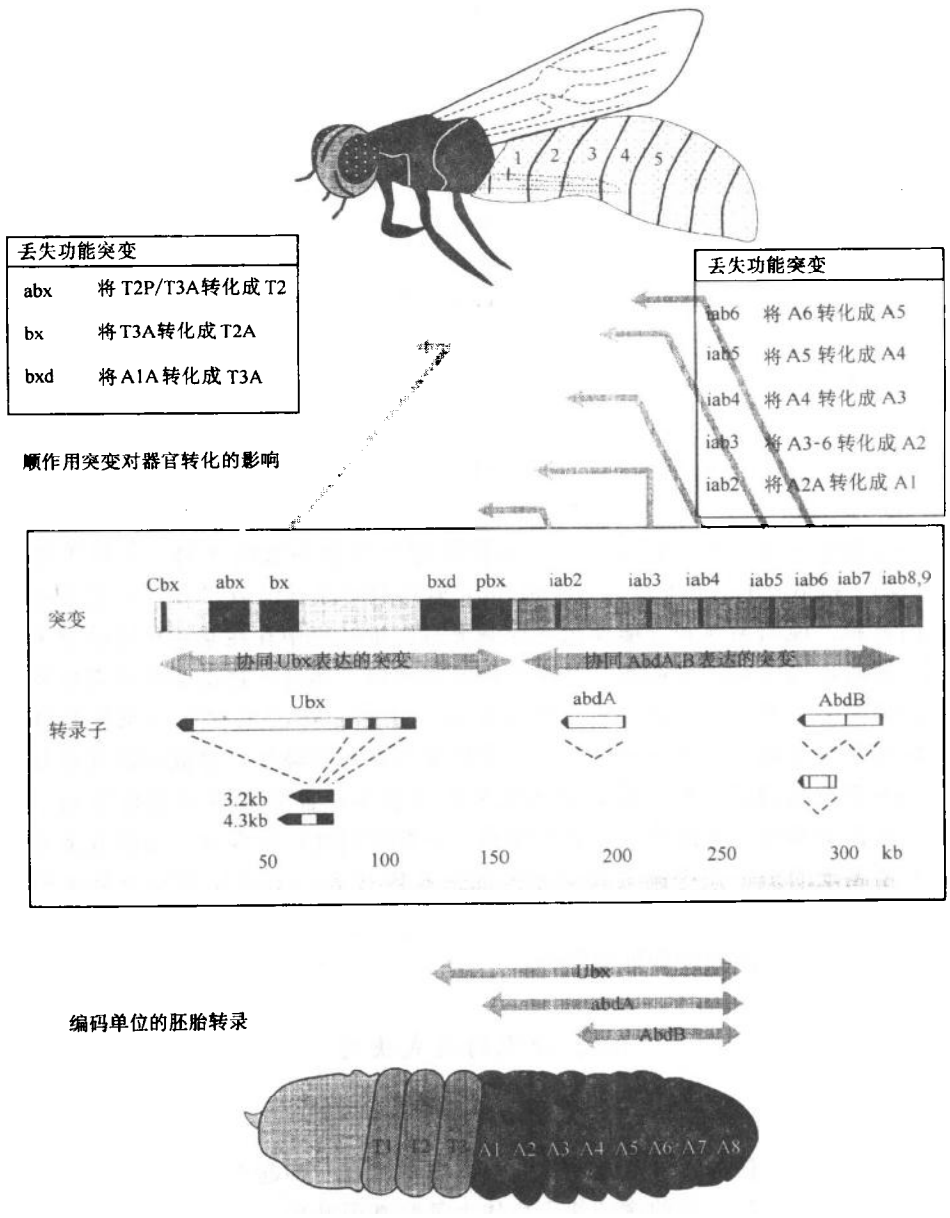


图 8－22　BX－C 座位的结构

同源异型基因表达受到在它们发挥作用之前的分节基因的调控，又受到相互之间的调控。因此，同源异型基因的表达受到复杂的时空调控。大多数同源异型基因是在有限的空间表达的，每个基因的表达空间相当于一个准体节，其

顺式调节序列和启动子相距甚远，要经由调节蛋白使 DNA 环化才能发挥作用（见第四章）。反式作用因子分为三群：第一群是分解基因的产物，沿前后轴按体节限界表达，起正负控制作用；第二群是阻抑蛋白，如 polycomb 群的 12 个基因；第三群是激活蛋白，在转录中、转录后和翻译后起正控制作用。同源异型基因的限界表达取决于空间信号，如 ubx、lab、dpp、abd－A 和 wg 等相邻基因表达之间的相互关系。许多同源异型基因编码的是转录调控因子，调控自身及相互之间的转录水平，并选择性地激活不同的实施基因，从而使组织细胞向特定方向分化。目前，对同源异型基因的靶基因尚知之甚少。

果蝇中控制体型形成的各个基因都含有高度保守的 180bp 的 DNA 序列，称为同源异型框（homeobox），这一序列位于许多同型基因和体节基因的 3′外显子中，如 BX－C 复合体的 ubx、abd－A 和 abd－B 区，ANT－C 复合体的 dfd、Scr、ftz 和 antp 基因。它编码羧基端的 60 个氨基酸，称为同源异型域（homeodomain）。在进化中同源异型域是保守性的，斑马、鱼、爪蟾、小鼠、人体基因组中都存在多拷贝的、与果蝇同源异型框相似的序列。非洲爪蟾的 x1hbox2 和小鼠 Hox1.1 基因与果蝇 antp 编码的同源异型域仅有一个氨基酸残基的差别，说明果蝇和脊椎动物在控制形态发生过程中存在某些相同的分子机制，同时也说明在进化期间，基因产物或基因的一部分产物是某种极端选择压作用的结果。某些研究还表明，酵母的 mat α1 和 mat α2 转录子与果蝇和脊椎动物同源异型框的部分序列也具有一定程度的序列同源性，可能同源异型基因与 mat 座位的功能相似，具有启动或关闭控制主要器官发育的遗传密码。可见，在由酵母到人类的所有真核生物中，同源异型框广泛存在。同源异型基因是目前发现的少数几个调节发育途径的主基因代表。据估计果蝇中最多约有 200 个调节基因，发现和研究这些基因的功能对探明个体发育的遗传控制以及进化机理将产生广泛而深远的影响。

四、分化的反式决定

（一）激素对基因表达的正控制

激素在生物体内含量甚微，但在个体发育过程中却起着重要的作用。激素的调节作用是通过一定的途径使染色体上某些基因开放，合成特异性蛋白质实现的。多肽类激素的受体分布在细胞的最外表面，通过第二信使影响细胞内功能，类固醇激素受体分布在细胞质内，类固醇激素受体复合物进入细胞核后直接和细胞核基因组相互作用。

昆虫的蜕皮激素（ecdysterone）对基因表达的控制是研究比较清楚的一

种。果蝇的成虫盘的分化受两种激素的共同调控，即 20－羟蜕皮激素和保幼素。蛹化过程中，咽侧体不再释放保幼激素，20－羟蜕皮激素引起成虫盘分化，蛹变为成虫。蜕皮激素是一种类固醇激素，这种激素决定昆虫从幼虫期成功地过渡到蛹期。

在早期幼虫阶段，果蝇唾液腺中多线染色体的疏松区型（puffing pattern）是稳定的，但是，在从幼虫期过渡到蛹期时，一旦蜕皮激素释放到昆虫的血淋巴（hemolymph）中后，原来的疏松开始消退，并迅速形成 100 个以上新的疏松区。每一疏松区都在准确的时间形成，并在一定间隔时间内保持活性（图 8－23）。早期的疏松在激素释放后 5 分钟内开始出现，这些疏松区可能是激素直接作用的结果，因为其形成不受蛋白质抑制剂的影响。相反，当蛋白质合成

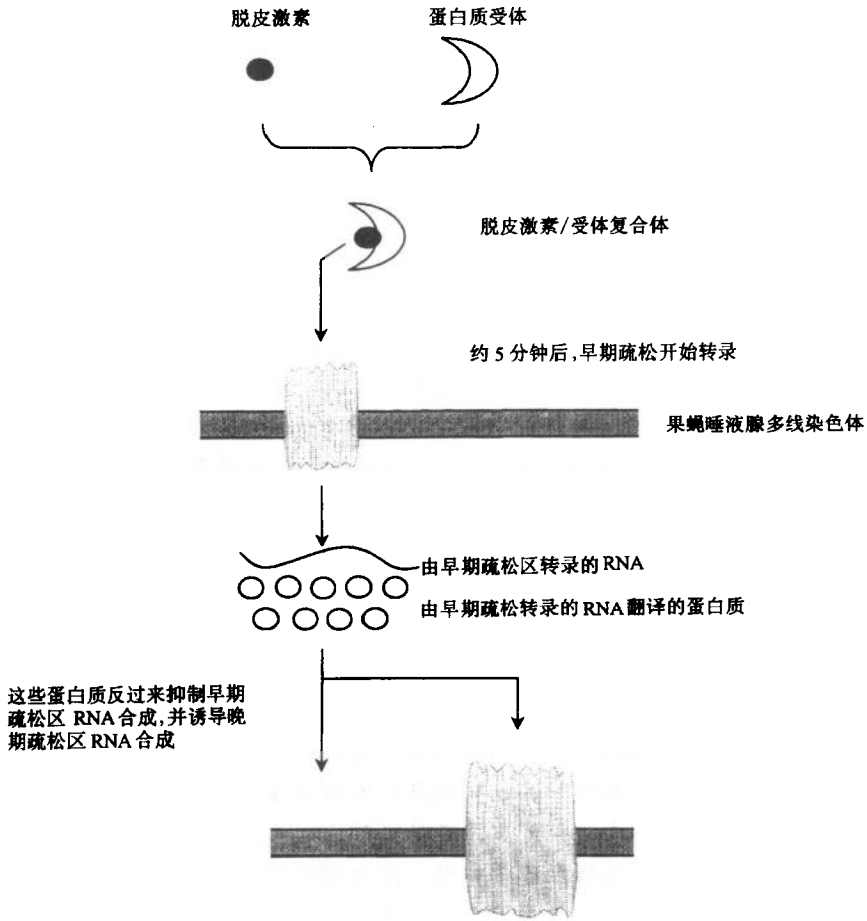


图 8－23　蜕皮激素诱导果蝇唾液腺染色体上疏松区形成的示意图

受抑制时，晚期疏松则不能形成。正常情况下，晚期疏松在加入蜕皮激素后数小时形成。可能由早期疏松区编码的某种蛋白质诱导晚期疏松区的形成。

蜕皮激素诱导早期疏松区迅速形成的现象表明，这种激素可能直接结合到染色质的某些特殊位点上。因此昆虫蜕皮激素的功能可能类似高等生物的类固醇的作用。蜕皮激素首先与某种特殊受体蛋白结合，然后这种激素/受体复合体再直接与特殊受体 DNA 序列结合，调节转录。虽然类固醇通常行使一种转录的正活化子的功能，但并不始终都是如此。蜕皮激素除了诱导疏松区扩展以外，它还使各种原有的疏松消退（再浓缩）。但是，似乎蜕皮激素直接与这些行将消退的疏松区结合，这与它诱导新疏松区形成的作用一样。这些现象表明，类固醇也能行使一种转录负调节因子的功能。因此，从理论上讲，一种转录的正活化蛋白通过识别某个基因的特殊 DNA 序列来活化转录，同时它也可能是另一基因的抑制子。如果活化蛋白行使抑制子的功能，那么第二个基因中的蛋白质结合位点必定位于对转录很重要的区域内。

（二）分化的花斑发育与调节发育

多细胞生物体是由多种类型的细胞构成的，不同类型的细胞构成了生物体的不同组织器官。这些组织器官都是由受精卵分裂产生的细胞后代增殖而来。细胞分化（cell differentiation）就是指个体发育中，细胞后代在形态、结构、功能上发生差异的过程。

细胞分化至少以两种不同方式进行，一种叫做花斑发育（mosaic development），也叫镶嵌型发育，另一种叫做调节发育（regulative development）。这两种分化方式基本反映了两种不同的分子生物学机制，其划分的根据是细胞决定的时间。所谓决定（determination）是指细胞已具有特殊的基因型的分化和特殊的发育潜能。花斑发育指预先决定的细胞分化，其胚胎在早期时细胞已决定，若花斑型胚胎在 4 细胞期时分离，每个细胞只能发育成 1/4 的胚胎，每个细胞的发育潜能是不可逆的。这种发育形式中，每个细胞的分化并不取决于相邻细胞的行为，每一个细胞内部事件指导该细胞做什么。这类事件可能由细胞自身的分裂所启动，或由内部生物钟的运转所启动。调节型发育指细胞保留着全能性，若把调节型胚胎在 4 细胞期使细胞分离，结果每个细胞都能形成一个完整的、体积较小的胚胎，每个胚胎都能发育成一个个体。

软体动物、环节动物和节肢动物的胚胎发育是花斑型的，这些动物卵所含形态决定子在卵裂过程中被分配到特殊的细胞内。最典型的镶嵌型卵的特征可用海鞘来说明。这种卵的色素比较复杂，而且当它处于单细胞的时候就显示出各种色素的不均匀分布。通过对正在发育的原肠胚中这些色素的分布情况，证明卵细胞质中一些特殊的区域将预定会发育成为胚胎的一些特殊结构。对其他

一些动物如线虫，蛔虫属和环节动物门的沙蚕研究也表明其发育也是花斑型的，如果把其中单个卵裂球从其周围的分裂球中分离出来，它可以通过细胞分裂达到成熟，并且只能长成胚胎中的一部分，相当于正常胚胎的预定图谱中所指明要发生的那部分组织。

海胆胚胎是无脊椎动物典型的调节型胚胎。在 2 或 4 细胞期时分离卵裂球，胚胎都能发育成完整的幼虫。两栖类中的瓦胚通常被认为是调节型发育，当两细胞期的卵裂球被分割，每个卵裂球都能发育成正常的胚胎，但每个卵裂球必须包含有一部分灰色新月区。哺乳动物的胚胎发育具有调节能力。在兔或鼠胚胎的两细胞期时，毁坏其中一个细胞然后将另一细胞移植到另一母体的子宫内，胚胎能正常发育。兔的胚胎在 8 细胞期时毁坏 7 个细胞，留下的卵裂球仍能正常发育。

在花斑发育和调节发育研究中的许多实验都支持了以下结论：花斑型卵能通过预先分化的卵细胞质明显地确定一个细胞的命运；高度调节型卵产生的胚胎，在原肠作用以后丧失其调节能力，以后它们的命运就会像花斑型胚胎的细胞那样受到限制；从原肠作用以前的调节型卵发生的胚胎中分离得到的细胞，如果单独培养可以发育成完整的个体，如果被移植到胚胎的另一部位，则能接受这个新位置上的“组织特点”。

值得指出的是，动物细胞的全能性随细胞分化程度的提高而受到限制，分化全能逐渐变窄，这是指整体细胞而言。细胞核则不同，它含有保持物种遗传性所需要的全套基因，而且并没有因细胞分化而丢失基因，因此高度分化的细胞核仍具有全能性。近年来，克隆动物的产生便是很好的证实实验。

五、细胞质定子

细胞核物质在每次卵裂时都均等地分配到各个子细胞，但受精卵细胞质中的物质分布及其在子细胞的分配不是十分均匀的，这种不均匀性对胚胎早期发育影响很大，在一定程度上决定着细胞的早期分化。细胞质定子（cytoplasmic determination）就是指细胞质中决定细胞核行为的物质。母体中的卵细胞不仅仅只是一种未分化的、待受精的单倍体细胞，它已具有复杂的内部结构，反映出卵细胞质广泛的区域性分化。细胞质定子以一种精确的、三维形式进行分布，在受精后它通过切割分裂被分配到不同的囊胚细胞中，最后控制卵裂球（blastomere）细胞的分化。

研究最清楚的细胞质定子是昆虫卵的后细胞质中的极粒（polar granule）。极粒使切割分裂的细胞核在其附近分化成原生殖细胞（primodial germ cell）。

受精之后，果蝇的合子核立即进行8次核分裂而不细胞化。经过9次分裂后大多数细胞核都迁移到卵的外围，然后再进行4次同步分裂，最后由卵膜将其包裹，形成细胞型的囊胚层。这些细胞进一步产生果蝇定子体组织（非生殖细胞）。但是，在经过8次核分裂之后，其中有少数核将分化成原生殖细胞，即在卵的后部，接近极粒的细胞核将分化成原生殖细胞。原生殖细胞又称极细胞（pole cell），比囊胚层细胞小而圆。极细胞开始位于囊胚层之外，但后来则迁移到性腺组织中，在性腺中产生数千个子生殖细胞。

根据对果蝇黄色隐性突变体的研究，由于只有一个正常基因的产物就能使果蝇形成正常体色，因此纯合体 Y^+/Y^+ 和杂合体 Y^+/Y^- 果蝇的颜色都正常，而纯合体 Y^-/Y 果蝇则呈淡黄色。如果用一支极小的吸管从野生型胚胎的后端移出极质，然后再移植到另一 Y^+/Y^+ 胚胎的前端，移植进来的极质便诱导形成大的圆形细胞，这种细胞类似第二个胚胎前端的极细胞，这种胚胎的前端通常不形成极细胞。为了验证极细胞是否具有遗传功能，然后再将极细胞移植到第三个基因型为 Y^-/Y^- 的胚胎中（这种胚胎将发育成黄色果蝇，因为其体组织都具有 Y^-/Y^- 遗传组成），但其后代中有些为野生型表型。这些野生型果蝇可能是从第二个胚胎 Y^+/Y^+ 中获得了一个显性基因 Y^+ 产物的缘故。这种迹象表明，极质在第二个胚胎中诱导了具有功能的极细胞。

脊椎动物中也存在极质。例如，蛙卵的植物极含有一种特殊的生殖质（germinal cytoplasm），植物极是形成原生殖细胞的场所。这种细胞质中的活性因子之一可能是核酸（DNA或RNA），因为在受精之前用紫外线有选择地照射生殖细胞质会导致不育。但是通过微量注射法将正常卵中的生殖细胞质注射到经紫外线照射的卵中，由这种卵发育成的蝌蚪具有发育完好的性腺和生殖细胞。在摇蚊（smittia）中，采用同样方式对其卵进行操作也获得相似的结果。在昆虫和两栖类动物中，生殖细胞分化的强烈相似性表明，在这些生物中存在高度保守的机制，保护原生殖细胞分化成体细胞。虽然对这种保护必要性的认识还不够，但可以认为，体细胞分化与DNA的物理修饰、DNA的体细胞重组或DNA的包裹发生变化等有关。还有一些实验表明，两栖动物卵中的灰月牙（gray crescent）中存在细胞质定子，在神经胚阶段（neurula stage）（囊胚之后），细胞质定子是肌肉特异性的肌动蛋白基因进行转录所必需的。

细胞质定子和形态发生梯度是决定胚细胞分化的两条根本不同的途径。细胞质定子只能指导与其临近的细胞核分化成特殊类型的细胞。相反，细胞依据其在梯度内的位置而作出不同的反应，定位信息的梯度能够影响很大范围内许多细胞。细胞质定子的优点可能是，当细胞不必花费时间来创造和翻译定位信息的梯度时，使胚胎发生以最快的速度进行。

六、植物花芽分化的同源异型基因

植物花的发育分为4个主要阶段：营养态向生殖态的转化、花序分生组织的形成、花器官的形成及其决定、花器官的生长和决定。有大量的基因参与调节花的发育，其中有些基因突变会引起一种结构发育成另一种结构，这类调节基因就是参与植物花芽分化的同源异型基因。目前这方面的成果主要是从拟南芥菜（*Arabidopsis*）的花发育研究中得到的。拟南芥菜的发育过程大致为：种子萌发，苗端分生组织产生螺旋状排列的叶，间以短的节间，使植株基部呈莲座状。营养生长期结束后，从分生组织产生花序分生组织并长出由长节间隔开的小茎生叶；之后出现花分生组织，每个花分生组织发育成一朵花。花的4种结构为花萼、花瓣、雄蕊和心皮成轮状排列，称1、2、3、4轮。

拟南芥菜中无花瓣（apetala）突变体形成同源异型花，即花瓣被抑制或被转变成其他器官。其中apetala－1基因的突变体不能形成花瓣，萼片（sepal）成为苞片（bract），成为无花瓣花，而且在这类花中萼片重复形成，所以形成花中花。拟南芥菜中野生型AP－2基因决定花被（perianth）器官的特征。根据对3个不同突变体（AP2－5、AP2－6和AP2－7）的研究，这3个突变基因都使萼片转变成心皮（carpel），AP2－5还使花瓣变成雄蕊（stamen），AP2－6使花瓣转变成心皮。

同源异型基因通常不只作用于一个花轮（floral whorl），而是作用于两个连续花轮，花由3个重叠区域组成，即萼片＋花瓣、花瓣＋雄蕊、雄蕊＋心皮，每一区域都受一套同源异型基因的影响。各个同源异型基因之间也可能进行互作，当然也与其他器官特化有关的目标基因进行互作。但目前对这些目标基因的特征及互作机制都不清楚。虽然认为同源异型基因具有定位效应，但其确切的作用是决定不同基因的临时表达。由于花轮的形成代表了分生组织上空间的连续性和时间的连续性，所以，以时间作标尺就可绘出同源异型基因的作用图，其作用可分成3个主要阶段，每一阶段都与一组以上的器官形成重叠(图8－24)。所以同源异型基因不仅是定位基因，而且也是定时基因，它们通过控制其他基因转录的时间而调节发育。

虽然同源异型基因的表达对体型形成是至关重要的，但目前还未分离到表达这些基因的细胞和确定发育的最早时期。为了形成花，分生组织必须经过发育相变化，而发育相变化必须依赖于每一花轮的生成细胞（founder cell）的活动。

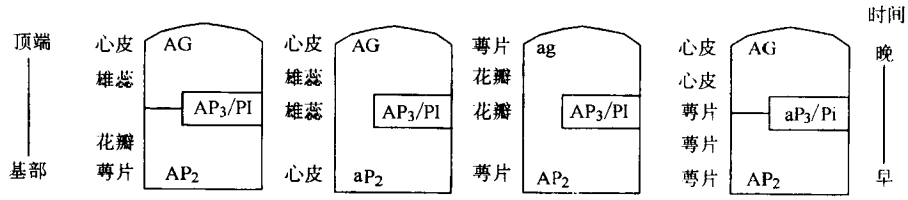


图 8－24　拟南芥菜中同源异型基因的互作

AP2、AP3/PI（Apetala3/Pistillata）和 AG（Agamous）基因可能作用于芽端 3 个重叠的空间部分的基部、中间区和顶端区。一般认为 AG 和 AP2 的作用在其领域是相互排斥的。如果其中之一为突变型，则另一个在其他各个区域都能发挥作用。空间顺序也代表时间顺序

在动物发育过程中，从特殊的胚胎细胞克隆可追踪到特殊的生成细胞。在海胆的发育过程中，这些细胞系的基础细胞占据了部分空间，它们在行使特殊功能时需要表达一些特殊基因。显然，营养体植株的分生组织与动物胚胎相比在发育上就具有更高的可逆性。顶端启始部分并不是永久性的，如其中几层构成原套－原体（tunica－corpus）的细胞层在叶启始形成时能够波动。但是最初的花器如萼片在多肽成分方面与叶片的相似性高于花瓣、雄蕊或心皮，但是，一旦分生组织转变成形成花时，其生长也就要终止了，并且这些分生组织中的启始早期花轮形成的空间区就可能与海胆胚胎的情况不一样。在海胆胚胎系统中，所有基因产物都被认为是细胞系特异基因表达的结果。这些基因可分为 4 类：第一类，当基础细胞群体形成后，这些基因很快被活化；第二类，衍生于特殊基础细胞群体的细胞系中被抑制的基因；第三类是看家基因，但这类基因在特殊基础细胞系中的表达水平随时间推移逐渐减弱；第四类是在特殊细胞系的发育过程中表达较晚的基因。

我们可以将决定花发育的同源异型基因归为第一类基因，决定叶特异表达的基因归为第二类，与光合作用有关的看家基因归为第三类，而将与色素形成和性别表达的基因归为第四类。第三、四两类基因的表达可能由第一类基因启动，而第二类基因的表达可能受第一类基因产物的控制，花形成的启动过程可能刺激顶端内不同部分，以促进同源异型基因的表达。刺激效应可使细胞建立一种协调网络系统，以一种有节奏的方式启始花器的形成。这一观点与每一花轮的生成细胞的临时活化相矛盾。当花形成的诱导过程受阻时，分生组织就保留在营养生长状态，或者花分生组织的局部转变成营养结构。根据上述模型，生成基础细胞中的第一类基因可能受诱导花反转的因子的抑制。受抑制程度可能反应在分生组织是否从启始花形成的花原基转变成营养体原基（急性抑制），或者分生组织是否转变成启始异常花器的形成，如类花瓣的雄蕊（温和性抑制）。当然这种模型需进一步证实。

第七节　基因对细胞周期的控制

细胞增殖（cell proliferation）是生命的重要特征之一。细胞通过细胞周期完成分裂，进行增殖，决定个体的生长发育和繁衍后代。细胞周期的运转是受到严格监控的，在酵母等低等真核生物中，目前已查明了多个与细胞周期调控有关的基因，细胞分裂的正常进行依赖细胞周期各个进程的基因产物功能的正常发挥。本节重点介绍参与细胞周期调控的基因及其作用。

一、细胞周期事件与基因表达

（一）细胞周期事件

细胞从上一次分裂结束到下一次分裂终了所经历的过程叫细胞周期（cell cycle）。一个周期所占的时间为细胞的一个世代（generation time），所以它在细胞水平上是时间的基本单位。

在20世纪50年代初，通过^{32}P标记蚕豆根尖细胞以及利用放射自显影技术对大肠杆菌的研究，提出了真核生物细胞一个典型的细胞周期是由4个不同的时相组成的，即有丝分裂期（M期），DNA合成期（S期）、M期结束到S期开始之前的间歇期（G1期）以及S末期到M期开始的另一间歇期（G2期）。正常情况下，细胞周期沿着G1－S－G2－M期的路线运转，通过M期一分为二，成为两个子细胞。

G1期：染色体为单线状，DNA尚未复制。但细胞内进行着一系列极为复杂的变化，为DNA的复制及蛋白质合成做准备。G1初期阶段积极合成RNA，其中包括mRNA、rRNA及tRNA。这些RNA在核内合成后，经核孔进入细胞质内，它们决定着合成具有特异性的蛋白质及酶。多肽在G1中期形成；G1后期合成DNA前体物质脱氧核糖核苷酸及胸苷激酶等。诱导DNA合成的物质也在G1期形成，动物细胞在G1末期中心体分离，为进行中心体复制做准备。

S期：DNA合成期，是细胞周期中最重要的一环，此时有关DNA合成的一些酶也相应出现活性高峰，组蛋白基因的表达也达到峰值，产生的组蛋白用于核小体的组装。S期内RNA的合成也很活跃，RNA量也增加一倍。RNA和RNA聚合酶与DNA的合成密切相关，若RNA合成受阻时，DNA复制也受阻。S期内，DNA进行准确复制，S期末，DNA含量增加一倍，保证了分裂后子细胞的遗传物质的平均分配。S期内染色质不同部分DNA复制先后不

一。DNA 首先在常染色质内进行复制，异染色质部分的 DNA 复制较晚，多在 S 期的后期进行复制。

G2 期：DNA 复制完成到有丝分裂开始的间隙时间，又称复制后期或合成后期。G2 有活跃的蛋白质合成，如微管蛋白等，晚期开始有丝分裂因子的合成，此期的生物合成是为 M 期的有丝分裂做准备。

M 期：有丝分裂期，细胞在 M 期进行有丝分裂。M 期又可分为前期（prophase）、中期（metaphase）、后期（anaphase）、末期（telophase）。此期 RNA 合成停止，蛋白质合成减少，染色体高度螺旋化，纺锤体形成，核膜破裂，染色体分裂并向两个极移动。M 期结束，细胞一分为二。

有些细胞还具有 G0 期（静止期），如高等动物的细胞可分为三类，即连续分裂细胞，休眠细胞，终端分化细胞。休眠细胞暂时脱离细胞周期，不进行增殖，但在适当刺激下，可重新进入细胞周期，如某些淋巴细胞、肝、肾细胞及大部分骨髓干细胞等，这类细胞都称作 G0 期细胞。胚胎发育早期所有细胞均为周期性细胞，以后随发育成熟，某些细胞进入 G0 期。植物细胞的 G0 期比动物及酵母都要明显。

细胞周期的有序运转是基因有序表达的结果，受到一些控制点的调控和监视。如细胞通过一个限定点在 G1 后期开始转录与 DNA 合成有关的基因，此限定点在酵母中称为“启动点”（start），哺乳动物中称为“限制点”（restriction point）。在开始合成 DNA 或进入 S 期之前还有一个“检查点”（check point），检查染色体 DNA 是否有损伤，如 DNA 有损伤先要进行修复，然后才能复制，以免遗传信息出错。此外还有 G2/M 检验点。这些检验点严格地监视着细胞周期事件的发生、发展过程是否严格按程序进行（图 8－25）。

S 期存在 S 期活化子，即一种蛋白质激酶，它属于活化有丝分裂的激酶的范畴。进入有丝分裂期取决于细胞内一种预先存在的、叫做 M 期激酶（M phase kinase）的蛋白质。根据对非洲爪蟾卵细胞发育的研究，M 期激酶又叫做成熟促进因子（maturation promoting factor，MPF），有两个亚基，一个为催化亚基 P34，它使目标蛋白质中的丝氨酸和苏氨酸磷酸化；另一个亚基是叫做细胞周期蛋白（cyclin）的 P45，是调节 M 期激酶与相应基质反应所必需的。细胞周期蛋白在间期持续合成，在有丝分裂期末降解，其降解与 M 期激酶失活有关。

所有真核生物都可能采用相似机制调节细胞周期。酵母为单细胞生物，有些种通过不对称出芽过程进行有丝分裂，但是控制细胞进入 S 期和进入分裂期的程序和蛋白质都与非洲爪蟾卵细胞发育的情况相似。而且在亲缘关系很远的物种如酵母、昆虫、哺乳动物中具有同源性的基因都执行相同的功能。

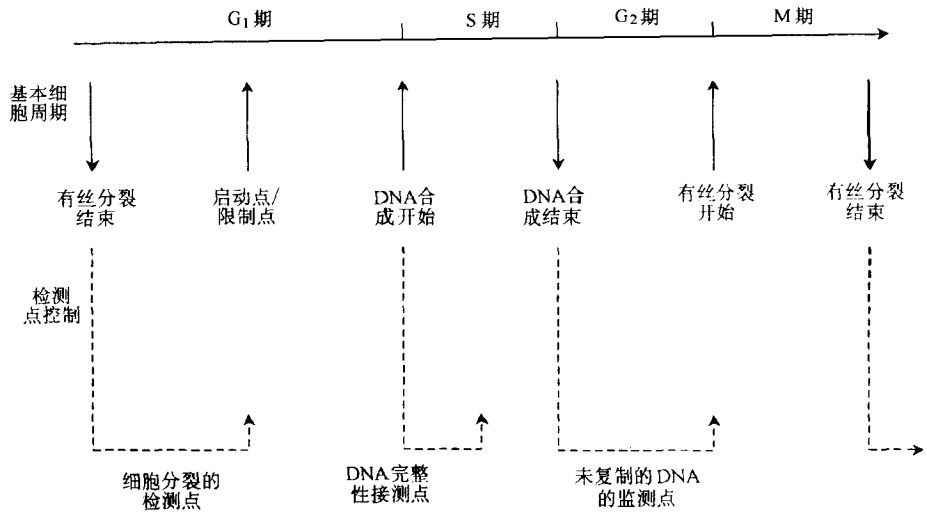


图 8－25　检测点通过检测细胞周期中前一个阶段是否完成来控制细胞是否进入下一个阶段

（二）细胞周期的调节因子

非洲爪蟾的 M 期激酶只有一个催化亚基 P34，但有多种不同的细胞周期蛋白亚基。细胞周期蛋白 A 和 B 是两种最普遍的形式，二者的相似性程度不高，约有 150 个氨基酸的序列同源，这段氨基酸序列有时也被称为细胞周期蛋白框（cyclin box）。在哺乳动物和蛙类中，细胞周期蛋白 B 又可分为两个亚型即 B1 和 B2。

因此，M 期激酶至少有两种普遍形式即 P34—细胞周期蛋白 A 和 P34—细胞周期蛋白 B。所有细胞周期蛋白的共同特征表明，它们都具有相同功能即影响 M 期激酶的催化亚基的活性。

在 G2/M 期活化 M 期激酶和在 M 期又使其失活的事件是确定细胞周期中各个检测点的关键。活化和失活是通过不同方式完成的。活化作用需要对 P34 催化亚基进行修饰，而这一反应是启动 G2/M 期转变的关键步骤。M 期激酶的失活是通过其细胞周期蛋白的降解实现的，一般 A 蛋白的降解发生在 B 蛋白降解之前。

是什么东西活化 M 期激酶的催化活性呢？P34 本身就是一种磷蛋白，其磷酸化状态是决定其活性的因素。为了使 M 期激酶具有活性，P34 某些位置的磷酸基团必须除掉，而另一些特定位置的磷酸基团必须保留。例如，家鼠细胞中，P34 的 ATP 结合位点上的两个磷酸基（第 14 位的苏氨酸和第 15 位的酪氨酸的磷酸基）必须被除掉，M 期激酶才具有活性。M 期激酶具有自动催

化特性，也就是说少量的 M 期激酶活化后可启动其他激酶分子的活化。使 P34 具有活性必须保留磷酸基的位置是第 161 位的苏氨酸的磷酸基。这一磷酸基团在 G2 期加上，在有丝分裂期末被除掉。所以，M 期激酶的活性是受其自身的磷酸化和脱磷酸化控制的。

细胞周期蛋白的降解是通过泛蛋白途径进行的（第四章）。这一过程与 M 期激酶失活有关，而且是细胞通过有丝分裂期所必需的。因此，M 期激酶失去其激酶活性是完成有丝分裂过程的前提。

由于 M 期激酶的活化是启动 M 期开始的第一步，而其失活又是正在分裂的细胞通过 M 期所必需的，所以由 M 期激酶所调节的事件是可逆的，也就是说，M 期激酶作用的基质的磷酸化是使分裂细胞形成纺锤体所必需的，而同一种基质的脱磷酸化则是使细胞回到间期结构所必需的。那么 M 期激酶作用的基质是什么？根据离体条件下的研究，这些基质主要有：H1 组蛋白（染色体浓缩所需）、核纤层蛋白（核膜解体所需）、核仁蛋白（与抑制核糖体合成有关）以及其他一些结构蛋白和酶蛋白。H1 组蛋白在细胞周期中的酸化是研究较多的方面，它在 S 期加进 2 个磷酸基，在细胞分裂期又加进 4 个磷酸基，这些磷酸化作用都由 M 期激酶提供。

对 M 期激酶作用的研究主要通过对酵母中与非洲爪蟾的 P34 蛋白基因相等的基因 cdc 基因（cell division cycle gene）的研究而进行的。在裂殖酵母中，P34 蛋白的基因叫做 cdc2，在芽殖酵母中则叫做 cdc28，其编码产物蛋白质则分别叫做 $P34^{cdc2}$ 和 $P34^{cdc28}$，有时也叫做 cdc2 和 cdc18。在这两种酵母中，已知 B 型细胞周期蛋白在 G2/M 期与 $P34^{cdc2}$ 或 $P34^{cdc28}$ 组装在一起。

P34 催化亚基尽管在酵母、蛙类和哺乳动物的进化过程中产生了一定程度的差异，但它们在结构和功能上都具有很高的保守性。如酵母与人体的 $P34^{cdc2}$ 有 63% 的序列相同，当将克隆的人体正常基因转移到裂殖酵母的 cdc2 缺失突变体基因组中后，人体基因可以补偿 cdc2 的缺失功能。尽管细胞周期蛋白的保守性没有催化亚基的高，但 $P34^{cdc2}$ 始终都与一个细胞周期蛋白亚基结合。

在裂殖酵母和芽殖酵母中，现分别鉴定出至少 80 个细胞分裂周期基因（cdc gene），但并非所有这些基因都与调节细胞周期有关。裂殖酵母中与细胞周期有关的基因如表 8－2。裂殖酵母的许多 cdc 突变体都可在 G2 和 M 期的临界期和 G1 期 START 点这两个阶段阻断细胞周期，表明它可能编码一种关键的调节因子。由于 cdc2 与非洲爪蟾的 P34 是同效基因，它可能是控制上述两个检测点的催化亚基。

cdc2 在细胞周期的不同阶段与不同细胞周期蛋白结合。在细胞分裂期与

其结合的亚基是 cdc13 基因的产物，二者结合产生类似动物的 P34/细胞周期蛋白 B 型二聚体的 M 期激酶。在 G1 期，cdc2 的产物与类 B 型细胞周期蛋白 cig2 结合。同动物细胞中 P34 蛋白一样，裂殖酵母 cdc2 催化亚基的活性也是由磷酸化控制的。根据细胞周期蛋白在细胞周期中与 cdc2 结合的差异，可将其分成两类，一类为 G1 期细胞周期蛋白（G1 cyclin），另一类为细胞分裂期（或 G2 期）细胞周期蛋白（mitotic cyclin）。G1 期和 S 期由 cdc2/G1cyclin 二聚体控制，即这种二聚体控制细胞从 G1 期过渡到 S 期。G2 期和分裂期则由 cdc2/分裂期 cyclin 二聚体控制，即这种二聚体控制细胞从 G2 期进入分裂期。CDC2/CDC13 二聚体中 CDC2 的脱磷酸化作用需要 CDC25，CDC25 并不具有典型的磷酸化酶的特性，但它是活化 M 期激酶过程中与关键的磷酸化作用有关的一种蛋白质，它把守检测点，活化 cdc2，以保证在 M 期活化以前完成 S 期。另外两个基因 wee1 和 mik1 通过使 CDC2 中 Tyr15 磷酸化而抑制 CDC2 的活性。进入 S 期需要活化 G1/S 型激酶（裂殖酵母为 CDC2/CIG2），同时也须使 M 期激酶（CDC2/CDC3）失活。细胞在通过了 START 点以后，进入 S 期还需活化另一基因 cdc18。当 cdc18 活化后，CDC2/CDC13 M 期激酶必须被抑制，当 M 期激酶被活化后，CDC18 必须失活才能避免启始另一轮 S 期。另一基因 rum1 的编码产物也是 M 期激酶的抑制子，该基因在 G1 到 G2 期之间表达，使 M 期激酶在这一期间保持失活状态，保证 S 期得以进行。因此，M 期激酶在促进分裂期进行的同时，也抑制 S 期，从而产生了一个内源检测点，即只有在控制前一个事件的因子被逆转之后才能启始下一个事件的进行。

M 期激酶在动物细胞和两种酵母细胞中的结构保守性以及这些细胞都需要形成相关的二聚体来调节 START 点的现象表明，动物细胞也可能采用相似机制控制细胞周期。动物同酵母细胞最大差别是细胞周期激酶具有更多的亚基类型，也不像酵母那样在 START 点和 G2/M 点使用相同的催化亚基。

酵母和哺乳动物细胞中在 G1/S 点和 G2/M 点激酶的亚基成分如表 8－2。同细胞周期蛋白家族可以根据其与 cdc2 互作的效率进行分类一样，催化亚基也可以据其与细胞周期蛋白互作的能力分成许多类群。动物细胞中与细胞周期蛋白结合的催化亚基和细胞周期蛋白都有多种变异类型，而且此时并不使用 cdc2 亚基。有两个 cdk 基因（cdk2 和 cdk4）编码的蛋白质与几种 G1 细胞周期蛋白结合形成二聚体。CDK2－G1 细胞周期蛋白二聚体调节 G1 期 S 期的转变，而 CDC2 细胞周期蛋白 A 或 B 二聚体调节 M 期完成。

表 8－2　构成 M 期和 G1 期激酶的亚基

物　种	G1 期		M 期	
	催化亚基	调节亚基	催化亚基	调节亚基
裂列殖酵母	CDC2	cig1，2（B 型）	CDC2	CDC13（B 型）
芽殖酵母	CDC28	CLN1－3	CDC28	CLB1－4（B 型）
哺乳动物/蛙类	Cdk（p33） CDK4	细胞周期蛋白 A、 D1、D2、D3、E	P34	细胞周期蛋白 A、 B1、B2

（三）cdc 基因对酵母细胞周期的调控

酵母细胞是研究真核生物细胞周期调控机制简单而有效的模型。从 1979 年第一个 cdc 基因在裂殖酵母中发现起，遗传学家一直利用酵母这种非常有利于遗传研究的材料来揭示真核生物细胞周期调控的复杂机制。酵母细胞周期研究的工作进展非常迅速，推动着哺乳动物和其他真核生物细胞周期的研究进展。酵母细胞周期的研究是采用裂殖酵母和芽殖酵母（啤酒酵母）为模型进行研究的。虽然这两种酵母在进化上差别很大，但控制细胞周期的基因在两者都是相同的。而且裂殖酵母 cdc2 突变株所丢失的功能可以由人的同源基因或芽殖酵母同源基因 cdc28 来补充，说明这些基因是高度保守的。

1．通过 start 点的调控　与一般真核生物细胞一样，酵母细胞周期由一次分裂结束到下一次分裂开始都要经历相同的变化阶段即 G1－S－G2－M。芽殖酵母在 G1 晚期有 start 点，它决定是否经历下一个细胞周期。芽殖酵母 start 点的完成以纺锤体极体（spindle pole body）SPB 进行复制为标志。裂殖酵母在 G1 晚期也有 start 点，但通过 start 点并没有进行 SPB 的复制。根据对芽殖酵母和裂殖酵母的遗传分析，通过 start 点需要 $p34^{cdc28}$ 和 $p34^{cdc2}$。在芽殖酵母 start 点起作用的是 CDC28 和一组特定的酵母 G1 周期蛋白（CLN1，CLN2 和 CLN3）。一旦 G1 期细胞生长达到了临界体积，就通过 CDC28－CLN3 复合物启动 CLN1 和 CLN2 基因的转录。CLN1 和 CLN2 的峰值正好在 start 之前。CLN1 和 CLN2 转录后关闭 B 型细胞周期蛋白的降解，启动 CDC28－细胞周期蛋白 B 的抑制蛋白 Sci1 的降解，关闭单倍体细胞对性激素反应的能力并引发芽形成所需的细胞质极化。这样细胞就通过了 start 点。裂殖酵母中有 3 种周期蛋白，其中 Cig2 是主要的在 start 点和 S 期起作用的周期蛋白，CDC2－Cig2 复合物在细胞达到通过 start 点所要求的体积时被激活发挥作用，使细胞通过 start 点。

2．S 期 DNA 复制的控制　start 的过渡对起始 DNA 合成和完成 S 期是非

常重要的，细胞通过 start 点进入 S 期，S 期中开始并完成 DNA 的复制。芽殖酵母中 DNA 复制的起始分两步：第一，复制前复合体 pre－RC（prereplicative complex）的形成。per－RC 包含起始点复合体 ORC（origin－recognition complex）、CDC6 和 MCM（minichromosome maintenance）家族蛋白。ORC 在整个细胞周期中始终结合在染色体上。CDC6 具有确保细胞在有丝分裂之前进行 DNA 复制的作用，缺乏 CDC6 时，分裂周期仍然进行，但不进行 DNA 的复制。芽殖酵母的 MCM 蛋白家族包括 6 种蛋白质，即 Mcm2、Mcm3、CDC46/Mcm5、CDC54/Mcm4、CDC47/Mcm7 和一种与裂殖酵母的 Mis5 蛋白相关的蛋白质，任何一种 MCM 蛋白基因发生错义突变都会导致 DNA 复制不能正常进行；第二，由 CDC28－Clb 和在 G1 晚期出现的具蛋白激酶活性的 CDC7－Dbf4 发动复制起点的启动。DNA 复制启动后，CDC6 和 MCM 家族蛋白逐渐被去除。pre－RC 变为 post－RC（post－replicative complex），post－RC 只带有 ORC。

3. G2 向 M 期过渡的调控　G2/M 转折点是决定细胞一分为二的控制点，该处的调控蛋白为 $p34^{cdc2}$ 和细胞周期蛋白 B 的复合物。芽殖酵母在 G2/M 转换处起作用的 $p34^{cdc28}$－细胞周期蛋白 B 复合物为 CDC28－Clb1、CDC28－Clb2、CDC28－Clb3 和 CDC28－Clb4。裂殖酵母在 G2/M 转换处起作用的 p34cdc2－细胞周期蛋白 B 复合物为 CDC2－Cig1 和 CDC2－CDC13。当细胞周期自 S 期进入 G2 期时，开始有细胞周期蛋白 B 的合成，细胞周期蛋白 B 是由 CDC13 编码的，合成后的细胞周期蛋白 B 存在于细胞质中，之后进入核内，细胞周期蛋白 B 与 CDC2 结合而被激活。细胞周期蛋白 B 的含量在 G2 向 M 期过渡时达到峰值，有丝分裂中期向后期转换时，细胞周期蛋白 B 逐渐被降解。$p34^{cdc2}$－细胞周期蛋白 B 复合体激酶活性在 G2 晚期出现，到 M 期活性最大，G1 早期活性消失。其活性的调节受到几个因素的控制：①细胞周期蛋白 B 与 $p34^{cdc2}$ 首先结合；② 裂殖酵母中 $p34^{cdc2}$ 的 Tyr15 的去磷酸化具有活化激酶的作用（芽殖酵母中不具备），催化这一过程的是 CDC25（Tyr 磷酸酶），而 Weel（由于基因突变导致小细胞分裂而命名）是 Tyr 激酶，使 Tyr15 磷酸化，因此 Weel 和 CDC25 活性间的平衡决定着 $p34^{cdc2}$－细胞周期蛋白 B 的活性；③ 细胞抑制因子 CSF（cytostatic factor）对 $p34^{cdc2}$－细胞周期蛋白 B 复合物活性起作用，原癌基因 c－mos 具有此种 CSF 活性；④$p34^{cdc2}$ 的 Thr167 的磷酸化是 $p34^{cdc2}$－细胞周期蛋白 B 复合物表现活性所必需的，裂殖酵母中催化 Thr167 磷酸化的是 Mop1；⑤ DNA 复制的结束也是 $p34^{cdc2}$－细胞周期蛋白 B 复合物活性变化需要的条件。$p34^{cdc2}$－细胞周期蛋白 B 复合体的活性下降后，细胞中一些有关磷酸化的酶开始发挥作用，使一些被 $p34^{cdc2}$－细胞周期蛋白 B 复合体催

化的底物去磷酸化，促使细胞由 M 期向 G1 期转化。

二、细胞生长的控制

（一）离体条件下调节细胞生长的因子

为了便于观察细胞结构功能及遗传性的变化和利用各种物理、化学或生物的外界因素来研究细胞生命活动的规律，在体外通过细胞培养获得性状相同的细胞作为研究对象来研究细胞生长的特点已十分普遍。高等真核生物的细胞处于一个复杂的“细胞社会”之中，除基本营养外，它们的生长还必须有各种生长因子的协调作用。广义的细胞因子包括生长因子、抑制因子（负性细胞生长因子，negative growth factor）、分化因子等，它们是体内外对细胞生长有调节作用的物质。正常培养的动物细胞，尽管培养基中有完全的氨基酸、维生素、无机盐，若不加入血清，细胞仍停止于 G0/G1 期。自从 Sato（1976）开始用无血清培养基研究细胞生长以来，细胞生长因子的研究得到了推动。至今已克隆的细胞因子已有四、五十种，且随着研究的纵深发展，其在细胞生长调节方面的重要作用也越来越受到重视。这里仅对高等动物几种常见的细胞生长因子进行介绍。

1. *神经生长因子*　神经生长因子（nerve growth factor，NGF）是最早发现的细胞生长因子，其发现始于对小鼠肉瘤的研究。1948 年 Bueker 发现 S180 小鼠肉瘤瘤块植入鼠胚中有神经纤维长入的现象。Cohen 和 Levi－Montalcini 用蛇毒磷酸二酯酶灭活肉瘤中的核酸，意外发现蛇毒中含有大量的神经生长活性物质，1958 年 Levi－Montacini 报道小鼠颌下腺也含有丰富的 NGF。1960 年 Cohen 报道由小鼠颌下腺取得纯化了的 NGF。人的神经生长因子基因定位于 1 号染色体上，有 6 个外显子，由两个相同肽链组成，每个肽链含 118 个氨基酸残基，单链分子量为 13KDa。

NGF 具有促进神经系统的生长发育和维持功能的作用，在大鼠的嗜铬细胞瘤细胞系（PC－12）培养液中添加微量的 NGF 后，PC－12 细胞呈现交感神经元的表型性状。另外，垂体内分泌腺部分、睾丸、附睾、卵巢等器官也有 NGF mRNA 的表达，说明 NGF 可能与这些器官的功能有关。

2. *表皮生长因子*　表皮生长因子（epidermal growth factor，EGF）是 1962 年 Cohen 等在纯化颌下腺 NGF 过程中，发现另有一类因子可引起小鼠眼睑早开、门牙早出、毛发生长受抑制等现象，将其命名为表皮生长因子。10 年后 EGF 的一级结构得以阐明，成为第一个被阐明结构的生长因子，即它是由 53 个氨基酸组成、分子量 6045KD、等电点为 4.6 的蛋白质。人类 EGF 基

因定位于4号染色体，长120Kb，包括24个外显子和23个内含子。人和小鼠的EGF有60%～70%同源性。

体外细胞培养实验表明，EGF对多种组织来源的上皮细胞都有促增殖活性，人和大鼠的成纤维细胞、人神经胶质细胞、人角化细胞、兔软骨细胞，人和牛的羊膜细胞及平滑肌细胞，牛、猪、人的卵巢颗粒细胞，大鼠肝细胞，牛的角膜内皮细胞，人的血管内皮细胞对EGF都敏感。EGF通过促进小分子物质如K^+、脱氧葡萄糖的输送作用，促进糖分解和乳酸增加和DNA、RNA及蛋白质合成，从而促进细胞增殖。

3. *成纤维细胞生长因子*　成纤维细胞生长因子（fibroglast growth factor，FGF）最早是1974年Gospodarowicz从牛脑垂体发现的对成纤维细胞有明显促进作用的因子。FGF广泛存在于多种组织中，按其等电点不同，分为碱性FGF（bFGF，pI=9.6）和酸性FGF（aFGF，pI=4.7）。人类aFGF基因定位于5号染色体，含有3个外显子，编码由155个氨基酸残基组成的蛋白质，分子量18KDa，bFGF定位于4号染色体，含3个外显子，指导若干种mRNA的翻译，编码不同分子量的蛋白质，以155个氨基酸残基的最为常见，分子量17.8KDa。两者有55%的氨基酸序列完全相同。

FGF的靶细胞均有相应的、特异的表面受体。FGF具有促进细胞有丝分裂的作用，它既能刺激成纤维细胞生长，又对血管生长有强刺激作用。FGF的体外作用很强，在pg/ml～ng/ml浓度范围内即可对人和牛的羊水成纤维细胞、牛的角膜上皮细胞及牛的晶体上皮细胞产生强烈的增殖促进作用。另外，FGF还与垂体细胞的生长和分化有关，还能促进颗粒细胞的生长，使卵泡成熟，并对胚胎的生长也具有促进作用。因此，FGF是一类高度多功能的细胞因子。

4. *转化生长因子－β*　转化生长因子（transorming growth factor，TGF）因其可使成纤维细胞转化为转化细胞而得名。哺乳动物的TGF－β至少有四种亚型，即TGF－β1、TGF－β2、TGF－β3、TGF－β1β2，鸡及两栖动物还分别存在TGF－β4、TGF－β5。人TGF－β1、TGF－β2、TGF－β3基因均定位于9号染色体上，均含有7个外显子，6个内含子，其核苷酸序列有高度同源性，分别编码相应的前体蛋白，TGF－β1含390个氨基酸残基，TGF－β2、TGF－β3含412个氨基酸残基。TGF－β1是同型二聚体，TGF－β2是由β2亚单位构成的同型二聚体，TGFβ1β2是异型二聚体，β1和β2氨基酸有70%同源性。人和小鼠的β1有99%的同源性。

TGFβ广泛存在于各种组织中，具有广泛的生物学作用，如可在表皮生长因子存在下刺激正常肾成纤维细胞在琼脂中生长。TGF－β一般对间充质细胞

起刺激作用，而对上皮或神经外胚层来源的细胞起抑制作用。

5. *胰岛素样生长因子*　胰岛素样生长因子（insulin－like growth factor，IGF）又称生长调节素（somatomedins，Sm），在体内主要存在于血液中，有IGF－Ⅰ和IGF－Ⅱ两种。人IGF－Ⅰ定位于12号染色体，IGF－Ⅱ定位于11号染色体。IGF－Ⅰ是碱性蛋白，由67个氨基酸残基组成，IGF－Ⅱ为近中性蛋白，由70个氨基酸残基组成，都含3个二硫键。IGF具有广泛的生物学作用，体外对软骨细胞、肌肉组织、成纤维细胞等组织细胞的培养实验表明，IGF可促进蛋白质合成和RNA合成，并促进DNA合成及细胞增殖。

（二）p53蛋白（p53protein）对癌细胞生长的负控制

p53是最重要的肿瘤抑制因子，在多种遗传或自发突变引起的人类恶性肿瘤中，p53基因突变是至今已知的与癌症相关的最普遍的基因变化，是了解肿瘤起源分子基础的研究热点。

1. *p53基因及其蛋白质的结构*　野生型p53基因全长约16～20Kb，不同物种的p53基因结构相似性很大，都有11个外显子和10个内含子。尽管人、鼠、爪蟾p53基因的内含子的长度有差异，但其外显子的部位有很高的同源性。早期的转染实验发现克隆的p53蛋白可使体外培养的细胞永生化，因此将p53基因划分为原癌基因。然而后来的研究证明，实用于转化实验的p53基因克隆实际上是p53的各种突变型，野生型p53蛋白具有抑制肿瘤细胞生长的作用，因而p53并非是核内癌基因，而是一种肿瘤抑制基因。野生型p53蛋白为核内磷酸化蛋白，分子量53KD，最初发现于由SV40转染的细胞中，与T抗原形成复合物。人p53蛋白全长393个氨基酸，小鼠390个氨基酸。许多转染细胞或肿瘤细胞中都可发现p53蛋白的大量存在。不同物种的p53蛋白都具有3个功能区和5个保守区域（图8－26）。3个功能区分别是：（1）带电荷的酸性N端区，包含前部75～80aa，其结构类似转录因子酸性区。（2）亲水的富脯氨酸区。人的为80～150aa、小鼠为75～150aa，其结构类似转录因子的富脯氨酸区。（3）带电荷的碱性C端区。人的为319～393aa，小鼠为276～390aa，能与DNA专一结合。人和小鼠有80％aa结构相同。5个保守结构域为Ⅰ～Ⅴ，对应的氨基酸序列分别是：13～19、117～142、171～181、234～258、270～286。在人癌细胞中发现Ⅱ～Ⅴ结构域是p53基因的突变热点。人的p53基因定位于17号染色体上，小鼠的p53基因定位于11号染色体上。不同物种的p53 mRNA长度为2～3Kb，3′非编码区长度从小鼠的800bp到非洲爪蟾的1800bp不等，人类的p53 mRNA还包含一段Alu序列。

p53基因突变对其蛋白质的性质有多方面的影响：（1）半衰期从20分钟增加到几个小时；（2）构象发生变化，某些野生型构象依赖的抗原决定簇的单

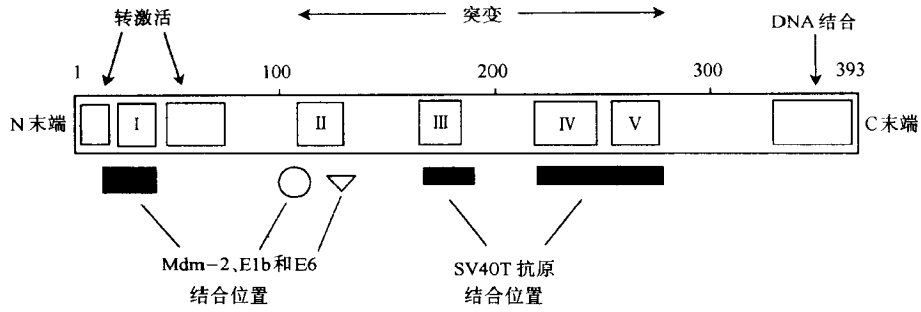


图 8-26　p53 蛋白的结构和功能区域

克隆抗体不能与突变型结合；(3) 合成后不能从细胞质进入细胞核。(4) 与病毒蛋白的结合很弱或不结合，而野生型的 p53 蛋白则与病毒的癌蛋白结合成复合物。

2. p53 蛋白对癌细胞生长的负控制　负控制 (negative control) 是指在没有调节蛋白时基因通常表达，加入调节蛋白时基因表达活性便被关闭。目前已有大量研究表明野生型的 p53 蛋白能负调节细胞生长，在一定条件下起到抑制肿瘤的作用。

正常细胞具有无限增殖的能力，但野生型 p53 蛋白的存在可使这种能力受到抑制。如果有影响野生型 p53 蛋白某个亚基功能的错义突变发生，则 p53 蛋白会失活，从而导致细胞无限制地增长，因此野生型 p53 蛋白被认为是肿瘤抑制因子。p53 基因突变在人类中引起一种叫做 Li－Fraumeni 综合征 (LFS)，即一种罕见的常染色体显性遗传病，其家族成员的一条 17 号染色体短臂缺失，个体发育和功能正常，但高发肿瘤，多数为乳腺癌、脑瘤、结肠癌和肉瘤。30 和 65 岁以下成员发病率分别为 50％和 90％，其原因是 p53 的另一等位基因在 245～258 位的氨基酸发生了点突变，完全丧失了生物功能。

正常的细胞均含有低水平的 p53 蛋白，p53 基因被激活的最常见原因是射线或其他因素导致的 DNA 损伤，此时细胞内有大量的 p53 蛋白表达。p53 基因被激活后可导致两种事件的发生：一是细胞生长受到抑制，一是细胞凋亡。细胞进入哪一种状态，多数情况下取决于细胞处于细胞周期的哪一个阶段（当然也有例外，如某些类型的细胞容易进入凋亡状态而不受细胞周期的影响）。如果细胞处在 G1 期，则细胞被阻断在 G1/S 检验点，不能进入 S 期。这时细胞会诱导 Cdk－细胞周期蛋白激酶 p21 基因活化，使细胞进入 S 期前 DNA 的损伤被修复，细胞可进入 S 期，细胞周期继续进行；如果细胞进入了分裂期，则 p53 蛋白引起细胞程序死亡或称凋亡。细胞凋亡的结果是染色质及细胞质浓缩，细胞体积缩小，后期 DNA 降解成寡核苷酸片段，这种凋亡体可及时地从

上皮细胞表面脱落或被吞噬细胞及时清除而不会破坏周围组织或细胞。

从理论上分析，上述两种途径的存在可能是由于DNA损伤会激活原癌基因，而p53蛋白则保护细胞避免这种现象发生。如果DNA损伤可被修复，则细胞可通过G1/S检验点，否则出现细胞凋亡现象。p53蛋白引起细胞进入两种途径的分子机制尚不清楚，但目前已知p53被激活后会发挥下列功能：

(1) p53蛋白可通过与靶基因DNA的特异序列结合而行使功能。与p53蛋白结合的区域都有共同的序列特征，为由0～13个碱基隔开的回文结构，每个结构单元由10个碱基组成，序列为5′－PuPuPu（A/T）（A/T）GpyPyPy－3′。p53蛋白的第120位赖氨酸、241位丝氨酸、273位及283位精氨酸，276位丙氨酸与p53靶DNA大沟的磷酸骨架接触，第120位赖氨酸、277位半胱氨酸，280位精氨酸与靶DNA的碱基通过氢键相互作用，第284位精氨酸与靶DNA的小沟通过多个氢键连接。第284位及273位精氨酸突变频率最高，可破坏p53蛋白与靶DNA的结合特性，使其失去转录因子功能。p53的90%以上的突变发生于此DNA结合区。

p53蛋白作为转录因子行使功能时，其中心部分与靶序列DNA结合，N－末端的反式激活域直接与TATA框结合蛋白相互作用。此区域也是一些其他蛋白的结合区域，如细胞癌蛋白MDM2和腺病毒E1B55KD蛋白，这两种蛋白均可结合在p53蛋白的N－末端而抑制p53蛋白的活性。如MDM2结合在N－末端会阻断反式激活域的作用而抑制p53的转录活化作用，其中p53蛋白的第19位苯丙氨酸、23位色氨酸、22位亮氨酸在与MDM2的结合中起重要作用。另外，p53蛋白的其他区域也可以被某些蛋白结合而使其功能受到抑制，如SV40大T抗原可结合在特定区域，阻断p53蛋白与靶基因DNA的结合。

(2) p53蛋白也可与不含上述靶序列的基因结合，抑制其功能，但其机制尚属未知。

(3) p53蛋白C－末端也具有与受损伤DNA结合的能力。这种结合是非特异性的，并通过识别单链DNA区域而实现。有人把纯化的C－末端肽（311～393位氨基酸）加到全长p53蛋白上，可成数倍增加p53蛋白与含GADD45基因（growth arrest and DNA damage gene）结合部位的寡核苷酸的结合能力，C－末端肽也激活p53蛋白与超螺旋DNA中GADD45结合部位的结合。缺失了C－末端肽的氨基酸（311～367位）则无此作用。

(4) p53的天然蛋白在溶液中为四聚体，第326～333位的氨基酸形成β片层，第335～354位氨基酸形成α螺旋，两者形成一个V型结构。上述区域有缺陷的p53蛋白在转染实验中不能抑制癌细胞生长。p53引起细胞凋亡的机制

尚不十分清楚。但 p53 的反式激活与其引起细胞凋亡这一功能之间似乎不存在相关性。腺病毒 E1B19KD 蛋白阻止 p53 引起细胞凋亡，但并不阻断其激活靶基因的活性，腺病毒 E1B55KD 蛋白阻止 p53 的反式激活能力，但并不影响其引起细胞凋亡的功能。因此，p53 引起细胞凋亡的途径可能可以不通过激活靶基因来实现。

另外有研究表明，bcl2 基因（B－lyphomo/leukemia－2 gene）与细胞凋亡和肿瘤发生有关，bcl2 是从 B 淋巴瘤中发现的凋亡抑制基因，它可使很多抑制凋亡的途径开启，加速细胞分裂、细胞增殖，维护细胞生存，延长细胞寿命，导致细胞数增加。大量实验证实，由 bcl－2 转染的细胞存活时间大大增长，这也证明细胞凋亡在抑制肿瘤方面有重要作用。野生型 p53 蛋白对 bcl－2 基因的表达有下调作用，这可能是 p53 蛋白引起细胞凋亡的途径之一。

可见，p53 对癌细胞生长的抑制作用的机理十分复杂，许多机制尚不清楚，但这方面的研究对于寻找恶性肿瘤的新基因治疗及免疫治疗方法都具有重要意义。肿瘤抑制基因为癌症治疗提供了一个广阔天地。随着研究的不断深入，癌变机制将逐渐被揭开，新的肿瘤防治对策也将出台造福于癌症病人。

三、细胞程序死亡

多细胞的真核生物在发育期间，细胞的死亡是发育过程的一种正常现象。一些多余的细胞在胚胎发生、形态建成及组织更新期间被清除掉，这一过程叫做细胞程序死亡（programmed cell death，或称 appoptosis），它是控制细胞总数的一种重要机制。在线虫（*C. elegans*）中，成体发育期间总数 1090 个体细胞中有 131 个细胞发生细胞程序死亡，即有些细胞在特定的时间和部位死亡。同样，在脊椎动物中，尽管各种体细胞系的形成尚不清楚，但在发育期间也会发生程序死亡，发生部位最明显的是免疫系统和神经系统。

细胞程序死亡需要活化细胞自杀途径，这一过程的典型标志是细胞变得紧凑、膜结构液泡化、染色质浓缩、DNA 断裂成片段。细胞程序死亡是一种主动过程，它依赖于行将死亡的细胞中 RNA 和蛋白质合成。很多因素都可诱导细胞程序死亡，如消除某些基本生长因子，用糖皮质激素处理，γ 射线处理，激活某些受体等。脊椎动物的免疫系统则采用另一种启动细胞程序死亡的机制，如胞毒 T 淋巴细胞攻击靶细胞等。肿瘤因子 p53 蛋白也可启动程序死亡，以防癌症发生。因此，细胞程序死亡不仅对组织的分化发育是重要的，而且对免疫防卫和鉴别癌细胞也具有重要作用。

细胞程序死亡的控制机制涉及一组抑制或活化这一途径的因子，例如，线

虫（*C.elegans*）基因 ced－3 和 ced－4 发生突变可使通常会死亡的细胞能够生存，表明这些基因与死亡有关。Ced－3 编码一种水解蛋白酶，ced－4 基因的功能尚不清楚，它可能属于另一系统的、功能与 ced－3 平行的基因。

Ced－9 基因抑制细胞程序死亡，使 ced－9 失活的突变是致死的，因为它使应该存活的细胞死亡。这一过程需要 ced－3 和 ced－4 参与，因此 ced－9 的功能是抑制 ced－3 和 ced－4 参与的过程。合理控制细胞程序死亡过程可能与这一途径的抑制和活化之间的平衡有关。

脊椎动物中与 ced－9 相当的基因叫做 bcl－2，它开始是作为一个原癌基因被发现的，因为在淋巴瘤中，易位可导致其过度表达。Bcl－2 蛋白的功能尚不清楚，其羧基末端固定到膜上，在线粒体和细胞核外以及内质网上都存在 Bcl－2 蛋白。由于这些膜位点都是氧原子生成的部位，表明 Bcl－2 具有抗氧化剂的活性，但这一功能与细胞程序死亡没有直接联系。

在鸡和人体中还发现了 bcl－2 的另外两个成员，分别叫做 bcl－X 和 bax。bcl－X 通过转录子的可变剪接产生两种特性不同的成分，即 bcl－Xl 和 bclXs。当 bcl－Xl 被转移到受体细胞中后，其作用类似 bcl－2，抑制细胞程序死亡；而 bcl－Xs 的作用则与 bcl－2 相反，使程序死亡得以进行。bax 基因的作用与 bcl－Xs 的作用相同。上述这些基因的产物在细胞中均以二聚体形式存在，可能 Bcl－2 纯二聚体使细胞不进入程序死亡途径，为细胞生存提供保护功能，而 Bcl－2/Bax 或 Bcl－2/Bcl－Xs 杂二聚体则不能提供保护功能，至于 Bax 和 bxl－Xs 纯二聚体的作用尚不清楚。总而言之，Bcl－2 蛋白质家族中的各个成员之间通过组合，形成对细胞程序死亡具有不同效应的二聚体。各个成员合成的相对比例是至关重要的，一个细胞是否进入程序死亡途径可能与 Bax 和 Bcl－2之间的比例有关。

抑制或活化细胞死亡的机制表明，可能所有细胞都具有内在进入程序死亡的功能。如果这一途径中的各种成分是普遍存在的，那么决定一个细胞死活的关键因素可能取决于这一途径被活化或被抑制的机制。

小　　结

生物发育分子机制的研究已成为生命科学领域的研究热点之一。对几种模型生物的大量研究表明，一系列发育事件的进行受阶段性表达的发育调控基因控制，其控制方式复杂而有序，即由多个基因群形成相互作用的层次网络，在时空上相互联系配合，导致个体发育的遗传程序不断展开。在形态发生的信号分子方面，Wnt 基因家族受到了广泛重视，在不同物种间或同一物种内均存

在多种 Wnt 同源异型基因，它们编码分泌型糖蛋白，在囊胚、原肠胚、中枢神经系统、器官发生及肢的发育等基本事件中起调控作用，其机制在小鼠、果蝇、爪蟾等动物中的研究较为深入。

λ噬菌体感染大肠杆菌后进入溶源途径还是裂解途径是研究基因控制发育的经典模型系统。cⅠ蛋白对许多（噬菌体基因的表达具有重要调控作用，它可使（噬菌体保持溶源途径的进行。cⅠ蛋白可与（噬菌体基因组中的操作子 OR 结合，结合力大小顺序为 $O_R1 > O_R2 > O_R3$。cⅠ与 O_R1、O_R2 的结合可产生两种效应：一是通过阻断 RNA 聚合酶与 cro 基因启动子的结合，直接抑制 Cro 蛋白的功能，同时在相反的方向激活控制 cⅠ基因转录的启动子而启动该基因转录。另一个在溶原途径和裂解途径转化方面起重要作用的蛋白是 CⅡ蛋白，CⅡ基因在λ噬菌体侵染早期表达，激活 cⅠ基因的转录，CⅡ蛋白的稳定性与宿主的营养状态有关。Cro 蛋白也可与 O_R1、O_R2、O_R3 结合，但其结合力大小顺序与 CⅠ蛋白正好相反。Cro 蛋白与 O_R3 结合可抑制 cⅠ表达，因此引起裂解途径发生。对λ噬菌体基因表达的研究表明，即使最简单的生物，基因也是以复杂的网络形式进行发育调控的。

细菌生长发育的各个阶段，基因表达也是按一定的时间顺序进行的。如枯草杆菌在不利条件下会产生芽孢，芽孢的形成过程涉及至少 6 个因子和约 100 个基因。鼠伤寒沙门氏菌的鞭毛蛋白基因的相变过程则涉及到 DNA 的重排，控制 H1、H2 基因交替表达，改变抗原性以逃脱寄主的进攻。

酵母交配型的转变的控制研究揭示了基因控制发育的更加复杂的层次网络。交配型的转变受 MAT 座位的控制，此座位编码调节特定交配型基因表达的转录因子，这些基因表达可使酵母以 3 种形式存在：α 和 a 单倍体及 α/a 二倍体细胞。与 MAT 在同一染色体上的还有两个交配型座位 HML 和 HMR，HML 含有不活动的 α 信息，它们与 α/a 交替转换有关，可转录出特殊的 mRNA，sir 基因的产物可调节其表达。交配型盖合运动的启始的控制较为复杂，ho 基因编码的内切酶使 Y、Z 片段之间形成缺口是启始过程的第一步，这种酶仅在双倍体母细胞 G1 期合成。交配型转换的方向由 α 或 a 细胞中 DNA 结合蛋白通过交配型座位进行直接或间接控制。α 和 a 细胞在融合成 α/a 二倍体细胞之前必须先转换成具交配感受性的细胞，这受到两种细胞释放的外激素的控制。

锥虫表面抗原转换的机制是研究原生动物发育的基因控制的典型例子。这种转换与酵母交配型的转换有些类似，但要复杂的多。锥虫在特定时间里只表达一种表面糖蛋白（抗原）。为躲避寄主的免疫系统的攻击，锥虫至少可以在 100 种不同的 VSG 间进行转换，这种转换通过表达位点或连锁拷贝的改变而

实现。

多细胞生物由许多具有不同形态特征和功能的细胞组成。动物发育的研究方面，果蝇胚胎发生过程中基因表达的研究是从分子遗传学角度阐明基因在多种层次网络上控制个体发育的理想模型。果蝇早期发育受4个基因群的控制，其中三个与前后极和末端系统的发育有关，一个与背腹极性系统的形成有关。前极系统负责头胸部的发育，Bicoid蛋白为此系统的形态发生素，它在胚胎前后轴方向上形成梯度，控制hunkback基因的转录，使其也在胚胎前后轴形成梯度。Hunkback蛋白决定着一些体节缺口基因（Krüpple，knirps和giant）的转录。母体基因的Nanos蛋白定位于早期胚胎的后极系统，Nanos蛋白梯度阻止hnukback mRNA在后极区的翻译。母体Bicoid基因产物与合子型体节缺口基因产物的相互作用控制着体节成对基因的转录，这些基因的表达将胚胎划分为14个发育单位即准体节。体节极性基因在每个准体节调节细胞发育成特定类型。最后，准体节中基因的选择性表达使身体不同部位的细胞长成不同的器官。末端系统负责胚胎两端一些特殊结构的发育，Torson受体蛋白在两极被特异激活。背腹极性系统负责背部发育，由母体toll mRNA编码的Toll蛋白可在背腹形成梯度。Toll活化的结果是建立起Dorsal蛋白在核中的活性梯度，dorsal基因是一个关键基因，Dorsal蛋白的不对称分布导致了中胚层或外胚层等结构的产生。同源异型基因是决定各体节形态特征的基因，这些基因在染色体上成簇排列。在果蝇中有两类这样的基因，ant－c和bx－c。同源异型基因中含有高度保守的DNA序列，叫同源异型框，在多种动物中都检测到了与此相似的序列，表明了体轴发育在进化过程中的高度保守性。植物中也存在同源异型基因，如控制花芽分化的基因，其中一种基因突变会引起一种结构发育成另一种结构。

激素对基因的表达具有正调控作用，例如果蝇的蜕皮激素可诱导多线染色体的疏松区迅速形成，决定了昆虫从幼虫到蛹期的转变。

多细胞生物体的细胞分化至少以花斑发育与调节发育两种形式进行，花斑型卵能通过预先形成的卵细胞质决定细胞的命运，调节型胚胎的细胞具有全能性。细胞中决定细胞行为的物质为细胞质定子，它控制卵裂球细胞的分化。

真核生物细胞的生长和分裂具有相似的调节机制，即通过细胞周期而实现细胞增殖。细胞周期是细胞从上一次分裂结束到下一次分裂终了所经历的历程，它沿着G1－S－G2－M期的路线运转。对酵母和爪蟾细胞周期的研究使我们对细胞周期的分子调控机制有了深入了解。细胞周期中两个关键调控点在于G1期和G2末期，G1的限制点（酵母中为start点）决定着DNA是否复制，通过此点后，细胞进入S期，此过程需活化G1/S期激酶（如裂殖酵母的

cdc2/cig2 及 cdc18 基因)；G2/M 期转折点是决定细胞一分为二的控制点，酵母在此点使用与 start 点相同的催化亚基；细胞进入有丝分裂期是由于 M 期激酶的存在，此酶为两个亚单位组成：即催化亚基 p34 和细胞周期蛋白 p45，所有真核细胞中均具有与之同源的蛋白存在。裂殖酵母中 p34 由 cdc2 编码，芽殖酵母中则由同源基因 cdc28 编码。M 期激酶的活性取决于其磷酸化状态，其活化是启动 M 期开始的第一步，由 cdc2 细胞周期蛋白 A 或 B 二聚体调节完成。M 期完成后，细胞中有关磷酸化酶开始发挥作用，此时 M 期激酶被抑制(如酵母中 p34cdc2－cyclin B 去磷酸化)，使 M 期向 G1 期转化。

机体各种细胞在生长过程中受到多种细胞生长因子的调控，这些因子在体外也能发挥作用，据此已克隆了几十种细胞因子，常见的有神经生长因子、表皮生长因子、成纤维细胞生长因子、转化生长因子及胰岛素样生长因子等。正常细胞均含有低水平的 p53 蛋白，它可使细胞生长受到抑制或引起细胞程序死亡，对该基因的研究对寻找恶性肿瘤新的治疗方法具有重要意义。细胞程序性死亡是多细胞生物生长发育期间的正常现象，它是一种主动过程，其控制机制涉及一组抑制或活化这一途径的因子，如线虫的 CED－3，CED－4，CED－9，脊椎动物的 Bcl－2，Bcl－X 和 Bax 等。细胞程序性死亡在组织分化、免疫防卫和鉴别癌细胞等方面具有重要作用。

复　习　题

1. 什么是发育定时钟 (DT)？以蛙卵受精后的发育为例说明 DT 作用的控制机制？

2. 举例说明基因控制发育的方式。

3. Wnt 基因家族与哪些发育现象有关？

4. λ 噬菌体侵染大肠杆菌后进入溶原还是裂解途径受哪些因子调控？如果 O_R1 位点发生了突变导致 CⅠ不能与之结合，Cro 蛋白的水平会起怎样的变化？此时带有这种突变的 λ 噬菌体会向溶原途径还是裂解途径发育？如果由于基因突变导致 Cro 蛋白不能与 O_R3 位点结合，cⅠ基因的表达会受到何种影响？此时 λ 噬菌体会朝哪一种途径发育？

5. 在 α 交配型的酵母中，控制 a 交配型的基因是如何被抑制的？同样，在 a 交配型的酵母中，α 控制交配型的基因是如何被抑制的？

6. 什么是鼠伤寒沙门氏菌的相变？说明其分子机制。

7. 锥虫是如何逃过寄主免疫系统的检测得以存活的？其抗原转变的途径有哪些？

8．什么是形态发生素？它是如何调节受精卵进行发育的？bicoid 和 nanos 的 mRNA 是在哪里合成的？bicoid mRNA 编码何种类型的蛋白质？如果一个突变体果蝇胚胎的前极仅含有低水平的 nanos mRNA，这对 Hunhback 蛋白的表达会产生什么影响？

9．果蝇卵的背腹极性是如何形成的？

10．什么是同源异型基因？其突变会造成怎样的后果？

11．什么是分化的花斑发育与调节发育？什么是细胞质定子？

12．检测点存在于细胞周期的哪个阶段？细胞在什么情况下才能通过检测点而继续进行有丝分裂？

13．M 期激酶的作用是什么？它是如何被激活的，又是如何失活的？

14．p53 蛋白被激活后可产生哪些功能？它发挥功能的可能途径有哪些？

15．什么是细胞程序性死亡？其被抑制或被活化的机制是什么？

第九章

基因操作技术及其应用

1972 年，美国斯坦福大学 P.Berg 领导的研究小组率先将动物病毒 SV40 的 DNA 与噬菌体 P22 的 DNA 连接在一起，第一次成功构建了体外重组 DNA 分子。1973 年，N.Cohen 等将几种不同的外源 DNA 插入 pSC101 中，并把这些体外重组质粒引入大肠杆菌中扩增，从而开创了重组 DNA 技术的研究。这一技术结合动植物细胞和组织培养、酶的固定化、现代化生物反应器和计算机技术的应用以及产品分离、纯化等技术形成了现代生物技术。它包括基因工程、细胞工程、酶工程和发酵工程四个主要方面。以重组 DNA 为核心的现代生物技术的创立和发展为生命科学注入了新的活力，它所提供的实验方法和手段极大地促进了传统生物学科如植物学、动物学、遗传学、生理学、生物医学等研究领域的深入发展。同时，也被广泛地用于医学、食品、化学、农业及环境保护等领域，为这些行业带来了一场新的技术革命。现代生物技术的发展仅有 20 多年，它在生命科学研究和产业化方面虽然已产生了巨大的影响，但这仅仅是个开始，生物技术的发展和应用方兴未艾。

第一节　DNA 重组的基本技术

重组 DNA 技术就是通过特殊酶的处理，使不同来源的 DNA 片段在体外发生重组，产生在自然界从未有过的重组体，然后将这些重组体分子转入适合的生物受体扩增或表达。重组 DNA 技术通常是从供体组织（donor）抽提并获得适合的包含一个或几个基因的 DNA 片段；将外源 DNA 片段插入到能够自我复制的载体（vector）DNA 分子中，该过程称为 DNA 重组（DNA recombination）；随后将重组 DNA 分子引入宿主细胞；重组质粒 DNA 分子一旦进入宿主细胞，载体将以正常的方式进行复制，同时插入的供体 DNA 序列也将随着载体一起复制。每一个重组的质粒在宿主细胞中将复制形成多拷贝。随着细胞的分裂，重组质粒 DNA 也将产生大量的拷贝。这种形成的单一的供体 DNA 片段的大量拷贝称之为 DNA 克隆。

重组 DNA 研究的基本方法，除了密度梯度超速离心和电子显微技术之外，还包括 DNA 分子的切割与连接、核酸分子杂交、凝胶电泳、细胞转化以及 DNA 序列结构分析和基因的人工合成等多种新技术和新方法。DNA 分子的切割与连接，无疑是基因操作的核心部分，而核酸限制性内切酶及连接酶则是关键的工具酶。

一、核酸酶和限制性核酸内切酶与 DNA 分子的体外切割

（一）核酸酶

基因的重组与分离须使用一系列有关的工具酶。现已发现了多种重要的核酸酶，其中有些可以降解 DNA，有些可以降解 RNA，有些可以切割单链核酸分子，有些则只降解双链核酸分子（包括 DNA 和 RNA）。例如能够在随机的位点使 DNA 分子内部发生断裂的内切核酸酶（endonucleases），可以从 DNA 分子末端移去核苷酸的外切核酸酶（exonucleases）等。

表 9－1　核酸酶的一般特性

类　型	特　性
核酸核酸酶	水解 RNA
脱氧核糖核酸酶	水解 DNA
内切核酸酶	切割内部磷酸二酯键
外切核酸酶	切割末端磷酸二酯键，或者从 5′端切割，或者从 3′端切割

某些 RNA 酶只水解邻近的那些含有特殊碱基的特殊的核苷酸的磷酸二酯键（表 9－2），这些酶对研究 RNA 的结构非常有用。还有一些 RNA 酶识别次级结构，对初级转录子进行加工。大肠杆菌的 RNA 酶 III 识别和切割 rRNA 初级转录子的双链主杆。

表 9－2　核糖核酸酶的基本特性

名　称	切割位点
Al	UpN 和 CpN
PhyM	UpN 和 ApN
T_1	GpN
U_2	ApN

在大肠杆菌中，至少有 6 种 DNA 外切核酸酶，这些酶都能从 DNA 的 3′端开始切割，其中有些酶也可以从 5′端进行切割；有些酶只能切割单链 DNA，还有些酶只切割双链 DNA。如外切核酸酶 III 能够切割双链 DNA，但是它也

能切割RNA－DNA杂种分子中的RNA，而不切割DNA。

还有一些核酸酶既能水解单链DNA，也能水解单链RNA，可广泛用于除掉DNA或RNA中未进行碱基配对的单链区，保留双链部分。最为熟知的是S1核酸酶，这种酶具有3′外切核酸酶活性，往往用于确定DNA分子上转录起点的位置（图9－1）。

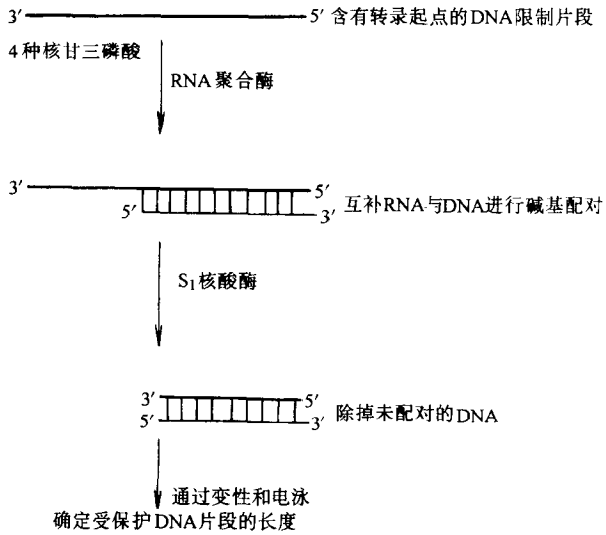


图9－1 S1核酸酶作图法，确定转录起点

（二）限制性内切核酸酶

限制性内切核酸酶（restriction enzymes）简称限制酶，最初是从原核生物中分离纯化出来、细菌用于抗噬菌体侵染的剪除和切割噬菌体DNA的特殊的内切核酸酶。到目前为止已经从300多种不同的微生物中分离出了约500种限制性内切核酸酶。内切核酸酶是指能在多核苷酸分子内部切割的酶，而限制性内切核酸酶则是指能够识别DNA分子特定序列，并在这一序列或在其附近切割DNA分子的核酸内切酶。

根据作用方式不同，限制性内切核酸酶又可分为三种，即Ⅰ型、Ⅱ型、Ⅲ型酶。这三种不同类型的限制酶具有不同的特性（表9－3）。其中Ⅱ型酶由于其内切核酸酶活性和甲基化活性是分开的，而且内切核酸酶作用又具有序列特异性，故目前在基因操作研究中应用广泛。

首先发现Ⅱ型内切核酸酶的是H.O.Smith和K.W.Wilcox（1970）。他们从流感噬血菌Rd株中分离出来的限制性酶具有典型的Ⅱ型内切核酸酶的特征。Ⅱ型酶识别双链DNA分子上特殊的靶序列，并破坏靶序列内的多核苷酸链。

表 9－3　限制性内切核酸酶的类型及其主要特性

	Ⅰ	Ⅱ	Ⅲ
1. 限制和修饰活性	单一多功能的酶	分开的核酸内切酶和甲基化酶	具有一种共同亚基的双功能酶
2. 限制内切核酸酶的蛋白结构	3 种不同的亚基	单一的成分	2 种不同的亚基
3. 限制作用所需的辅助因子	ATP，Mg^{2+}，S－腺苷甲硫氨酸	ATP，Mg^{2+}，（S－腺苷甲硫氨酸）	
4. 寄主特异性位点序列	EcoB：TGAN8TGCT EcoK：AACN6GTGC	旋转对称	EcoP1：AGACC EcoP15：CAGCAG
5. 切割位点	在距寄主特异性位点至 1kb 的地方可能随机的切割	位于寄主特异性位点或其附近	在距寄主特异性位点 3′－端 24～26bp 处
6. 酶催转换	不能	能	能
7. DNA 易位作用	能	不能	不能
8. 甲基化作用的位点	寄主特异性的位点	寄主特异性位点	寄主特异性位点
9. 识别未甲基化的序列进行核酸内切酶切割	能	能	能
10. 序列特异的切割	不是	是	是
11. 在 DNA 克隆中的用处	无用	十分有用	有用

注：N＝任何一种核苷酸。

常用限制性内切酶的识别序列一般为 4～6 个碱基，但也有 8 个碱基的（表 9－4）。切割位点的 DNA 的两条单链上识别序列从 5′端阅读时是完全相同的，这种结构叫做回文对称结构（palindrome）。切割位点通常位于识别序列之内，有些酶在两条识别序列的中间切割，形成所谓的齐头末端（blunt end）（例如 HaeIII）。但大多数酶都交错切割识别位点，形成一段单链尾巴，这种结构叫做黏性末端（sticky end）（如 EcoRI）。利用黏性末端可以将两条来源不同的 DNA 片段连接在一起，因为两条单链尾巴可以通过碱基互补配对粘连在一起。

表 9－4　部分限制性内切核酸酶的识别序列和切割位点

内切酶	识别序列*	内切酶	识别序列*	内切酶	识别序列*
Alu I	AG↓CT	Fnu4H I	GC↓NGC	BamH I	G↓GATCC
Hae III	GG↓CC	Hinf I	G↓ANTC	Bgl II	A↓GATCT
Hpa II	C↓CGG	Mae III	↓GTNAC	EcoR I	G↓AATTC
Mbo I	↓GATC			EcoR V	GAT↓ATC
Mse I	T↓TAA			Hind III	A↓AGCTT
				Sph I	GCATG↓C

* 识别序列的方向是 5′→3′，用↓表示酶的切割位点。因限制酶识别位点的两条链从 5′→3′读序列是一致的，切割位点也相同，故一般只列出一条链的序列。

如果 DNA 分子中碱基对的排列是随机的，则可用下列公式估算限制酶的识别序列在 DNA 分子上出现的频率：

识别序列的频率 = $(1/4)^N$ （N 代表识别序列的碱基数目）

由此可见，识别序列碱基数不同的酶切割 DNA 分子的频率差异很大。对识别序列为 4 个碱基的酶，其切点出现的频率为 $(1/4)^4 = 1/256$，也就是说在 DNA 分子上平均每 256 个碱基就有一个切点。但对识别序列为 6 个碱基的酶，其切点出现的频率为 $(1/4)^6 = 1/4096$。这样，我们就可以估算在已知长度的 DNA 分子上某种限制性内切酶切点的数目：

切点数目 = 分子大小（碱基对数）×识别序列频率

事实上，由于 DNA 分子碱基的排列并不是随机的，所以有时实际酶切位点和理论计算值有很大差距。如噬体 T7 的染色体组长 40kb，按公式计算，它应该分别有 10 个 EcoRI、Hind Ⅲ 和 BamHI 的识别位点，但实际上一个也没有。

当 DNA 用某种限制酶处理以后，就能产生一定数目的限制性片段（restriction fragment）。各种不同长度的限制性片段通过琼脂糖凝胶电泳很容易分离开来。DNA 片段在凝胶上运动的速度取决于其长度，所以片段越长，运动的距离就越短。电泳结束后，通过溴化乙锭染色，就能在紫外光下分辨各种不同的片段。如果有已知分子量的 DNA 作标记，通过比较，就可以确定各种片段的长度。

二、限制性片段长度多态性

如果将人体总染色体 DNA 用某种限制酶进行完全消化，就可产生大量限制性片段，通过凝胶电泳将这些片段分离开来后，利用基因组中目标片段作为探针，并通过 DNA 分子杂交，就可以探明各别具有同源性的片段。例如，人体基因组总体 DNA 用 HindIII 消化之后，再以 β－珠蛋白基因束的 γ 基因作探针，测定到长度为 2.7、3.5、7.2 和 8.0kb 的限制性片段与探针具有同源性。利用此法分析镰形细胞贫血症病人的 DNA 时，发现其中缺少 2.7 和 7.2kb 的限制性片段。造成这种现象的原因是这种病人的基因组丢失了限制性内切酶 HindIII 的两个酶切位点。上述这种现象叫做限制性片段长度多态性（restriction fragment lenth polymorphism，RFLP）。RFLP 是基因组中发生点突变的结果，而点突变则使基因组获得或丢失某种限制酶的作用位点。下图是假设的点突变造成基因组获得 HindIII 限制位点的过程。

$$\begin{array}{ccccc} AAGTTT & \xrightarrow{\text{点突变}} & AAGCTT & \xrightarrow{\text{点突变}} & AGGCTT \\ TTCAAA & & TTCGAA & & TCCGAA \end{array}$$

此外，缺失和插入也可改变两个限制性位点之间的片段长度。RFLP 现已用于某些人体遗传疾病的诊断和鉴别许多相近物种之间的遗传差异。

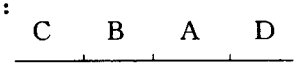
三、限制片段作图

如果一条很大的 DNA 片段用某些限制酶完全消化，就会形成几条较小的片段，其长度就可根据凝胶电泳后，各种小片段的长度来加以计算。如果再用同一种酶对原始 DNA 进行部分消化，就可以产生某些较大的片段，根据这些大小片段的关系推导完全消化后所产生的各种小片段在原始 DNA 中的顺序，这种作图方式叫做限制性作图（restriction mapping）。现在有多种进行限制性片段作图的方法，但通常作图时须采用两种以上方法，才能获得比较精确的图谱。现介绍其中两种作图方法。

（一）部分消化法

部分消化法是指首先以一种限制酶对 DNA 进行完全消化，然后再进行部分消化，最后根据两类消化中大小片段的关系，找出各种片段的相邻的片段，排出它们的顺序。如果某种限制酶在一条线状 DNA 分子上有几个切点，用此酶对 DNA 进行完全消化后，就可产生 n+1 条片段（环状分子为 n 条片段）。部分消化时，如果反应液中只有一条 DNA 分子，切割的次数和产生的片段数都应小于 n+1，但实际上，反应液中远不止一条 DNA 分子，所以部分消化的片段数总是大于完全消化的片段数。线状 DNA 经部分消化后，产生的片段数可用下式计算：$F_{n+1}=\frac{n^2+3n}{2}$，其中包括单一片段，二联片段、三联片段等等。

如果某种限制酶在一种线状 DNA 分子上有 3 个切点，部分消化时，最高片段数为 9，包括 A、B、C、D 单一片段，AB、CB、AD 二联片段以及 CBA 和 BAD 三联片段，如下图：



例如，从细菌中分离到一条 18.1kb 的线状 DNA，用限制 AvaII 完全消化，得到 8 条片段，部分消化则得到 14 条片段，求其图谱（图 9-2）。

完全消化的片段为：A（4.6）、B（4.0）、C（3.3）、D（2.5）、E（2.0）、F（1.0）、G（0.5）、H（0.2）

部分消化的片段为：6.6、6.5、5.3、4.6、4.2、4.0、3.8、3.5、3.3、

2.5、1.0、0.7、0.2

分析时，首先排除部分消化中与完全消化片段长度相同的片段，然后再分析剩下的片段。此例剩下片段有7条，即6.6、6.5、5.3、4.2、3.8、3.5和0.7。

由于大片段是完全消化的几条小片段长度之和，故6.6=2+4.6（A+E）3.8=0.5+3.3（G+C） 6.5=2.5+4.0（B+D） 0.7=0.2+0.5（H+G）5.3=2+3.3（E+C）

片段4.2有两种可能的组合，即4.2=0.2+4（H+B）=0.2+0.5+1.0+2.5（H+G+F+D）

片段3.5有三种可能的组合，即3.5=1.0+2.5（F+D）=0.5+1.0+2.0（G+F+E）=0.2+3.3（H+C）

以上述结果为基础，再根据有关片段之间的交错重叠现象，即可组建出线状DNA限制性片段的物理图谱。用完全消化的小片段拼组部分消化的大片段时，片段4.2有两种可能，片段3.5有三种可能，这里以其中前者的第二种，后者的第三种为可能的答案。

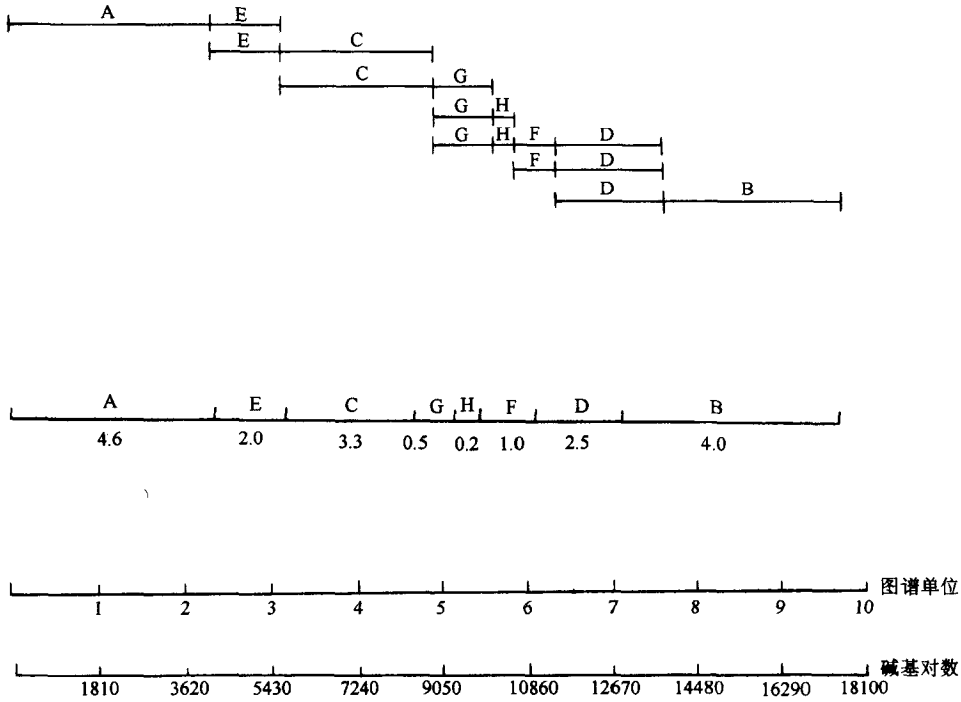


图9-2 18.1kb DNA的AvaII物理图谱

上为组排各条限制性片段的程序

（二）双酶消化法

双酶消化法是指在两种限制酶分别单独消化 DNA 的基础上，然后再用这两种限制酶同时或先后消化此 DNA 组建物理图谱的方法。双酶消化法与部分消化法的原理基本相同。

单酶消化后的某些片段在经第二种酶的消化后可能仍然保留，另一些片段则可能消失而出现新的片段，这种现象说明一种限制酶的切割位点正好位于另一酶切片段之内，而且新出现的有关片段之和必然等于某条消失的片段的长度。因此，根据片段的长度，我们就能找出消失片段与新生片段的关系，再根据它们的交错重叠现象，确定各条片段的顺序，组建 DNA 物理图谱。

例如，大肠杆菌的一个 10kb 的质粒 DNA 用 4 种限制酶进行单酶切和双酶消化，其结果如表 9－5。

表 9－5　单酶和双酶消化大肠杆菌 10kb 质粒的结果

消化类型	酶的组合	片段数	片段长度（kb）
单酶消化	EcoR I	1	10.0
	Hinc II	2	8.0，2.0
	Hpa I	1	10.0
	Sma I	2	5.5，4.5
双酶消化	Sma I＋Hinc II	4	4.5，2.5，2.0，1.0
	Sma I＋ EcoR I	3	5.5，2.5，2.0
	Sma I＋ Hpa I	3	4.5，3.0，2.5
	EcoR I＋ Hinc II	3	5.0，3.0，2.0
	EcoR I＋ Hpa I	2	5.0，5.0

根据片段大小，分别理清每一对酶的消化结果，除去单酶和双酶消化时产生的相同片段，再按照片段长度，找出新生小片段和一个消失的大片段之间的关系。如

Sma I＋ Hinc II，

Hinc II：8＝4.5＋2.5＋1

Sma I：5.5＝2.5＋2.0＋1

Sma I＋ Hpa I，

Hpa I：10＝4.5＋3＋2.5

Sma I：5.5＝3.0＋2.5

Sma I＋ EcoR I，

EcoR I：10＝5＋3＋2

Sma I：4.5＝2.5＋2

EcoR I＋ Hinc II，

Hinc II：8＝5＋3

EcoR I：10＝5＋3＋2

EcoR I＋ Hpa I，

Hpa I：10＝5＋5

EcoR I：10＝5＋5

根据上述结果，找出各条片段的相邻片段，组建此 DNA 的限制性图谱（图9－3）。

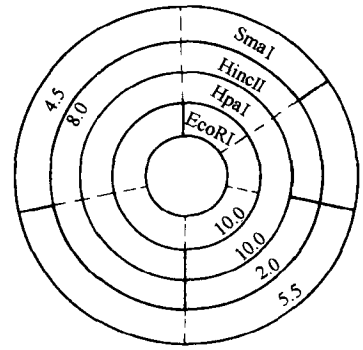


图 9－3　10.0kb 质粒的物理图谱

四、限制性片段的连接与克隆

如前所述，许多限制酶能够交错切割 DNA，在限制性片段的两端留下互补的单链尾巴，因此各条片段又可通过碱基配对而简单地连接在一起。但这种连接并不牢固，如果在反应液中加入连接酶和 ATP，这些片段就能永久性地连接在一起（图 9－4A）。连接酶还能催化具齐头末端的 DNA 片段的连接过程，但这种随机连接的频率很低。在 DNA 重组过程中，通常须将某种含有限制酶切点的接头连接到具齐头末端的限制性片段上，然后再用第二种限制酶切割，创造黏性末端（图 9－4B）。

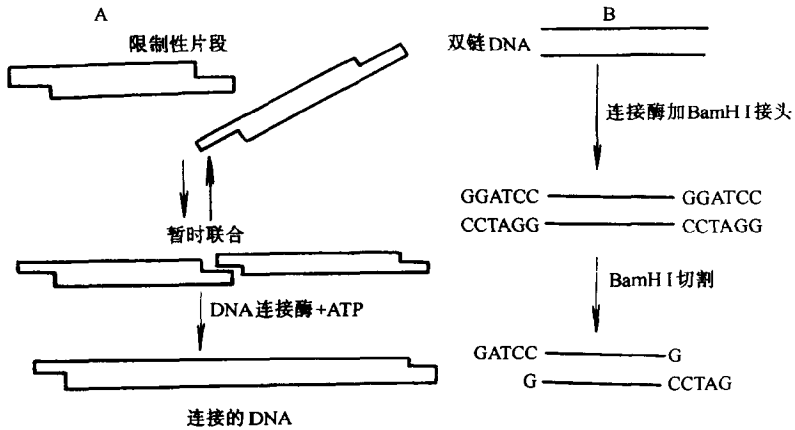


图 9－4　限制性片段的酶链

A. 在存在 ATP 或 NAD^+ 时，DNA 连接酶可将限制性片段共价连接在一起

B. BamHI接头加到双链 DNA 的齐头末端，以创造互补的单链尾巴

由于任何两条具有互补尾巴的限制性片段都能进行碱基配对，所以利用连接酶以及交错切割 DNA 的限制酶，我们就能在试管中将来自不同 DNA 分子的 DNA 片段重组在一起。

如果将目标 DNA 重组到某种能够自主复制的 DNA 分子上（图 9－5），然后通过转化（transformation）再将这种重组 DNA 引入适当的大肠杆菌，进行扩增，我们就能获得大量相同的细菌菌落。这个过程叫做基因克隆（gene cloning），这种能够自动复制的 DNA 分子就叫做克隆载体（cloning vector）。目前所使用的克隆载体通常为一些能自主复制的质粒和某些病毒 DNA 及其衍生物。

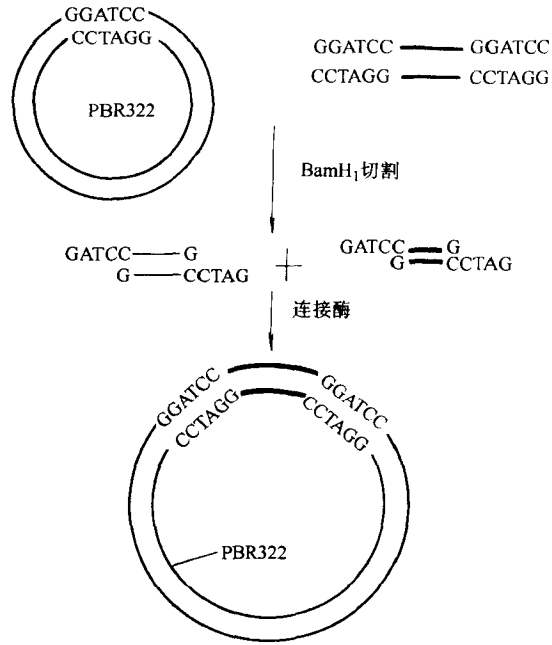


图 9－5　限制性片段插入到克隆载体上的过程

（一）DNA 分子的体外连接

1967 年，世界上有几个实验室几乎同时发现了一种能够催化在 2 条 DNA 链间形成磷酸二酯键的酶，即 DNA 连接酶（ligase）。这种酶需要 DNA 链的 3′末端有一游离的羟基（－OH）和在另一条 DNA 链的 5′末端具有一个磷酸基团，只有在这种情况下，才能发挥其连接 DNA 分子的功能（图 9－6）。由于在羟基和磷酸基团之间形成磷酸二酯键是一种吸能反应，因此还需要有一种能源分子才能实现这种连接反应。DNA 连接酶不能连接两条单链的 DNA 分子，被连接的 DNA 链必须是双螺旋的一部分。实际上，DNA 连接酶是封闭双螺旋 DNA 骨架上的缺口（nick），而不是封闭裂口（gap）。用于共价连接 DNA 限制片断的连接酶有两种不同的来源，一种是由大肠杆菌染色体编码的 DNA 连接酶，另一种是由大肠杆菌 T4 噬菌体 DNA 编码的 T4 DNA 连接酶。这两种 DNA 连接酶除了前者用 NAD^+ 作能源辅助因子，后者用 ATP 作能源辅助因子外，其他作用机理并没有什么差别。

DNA 连接酶最突出的特点是，它能够催化外源 DNA 和载体分子之间发生连接作用，形成重组的 DNA 分子。应用 DNA 连接酶这种特性，可以将具有

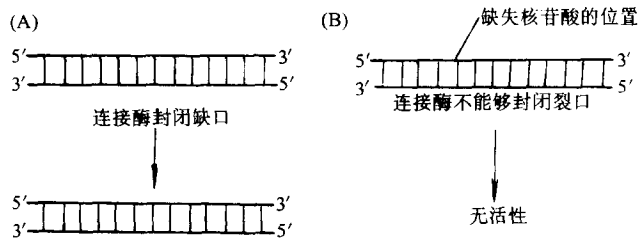


图 9－6　DNA 连接酶的活性

A. 具有 3′－OH 和 5′－P 基团的一个缺口被 DNA 连接酶封闭起来

B. 如果为缺失一个或数个核苷酸的裂口，DNA 连接酶则不能将它封闭

黏性末端的 DNA 限制片断插入到适当的载体分子上，从而按照人们的意图构建出新的 DNA 杂种分子。具黏性末端的 DNA 片断的连接比较容易，也比较常用。

重组 DNA 实验的一般程序是，选用一种对载体 DNA 只具有惟一限制位点的限制酶切割载体。例如 EcoRI 限制酶对小质粒 pSC101 DNA 就只具有一个限制位点，因此，经此酶消化之后就会形成全长的具黏性末端的线性 DNA 分子。再将外源 DNA 大片段也用 EcoRI 限制酶作同样的消化，就可以形成能插入到 pSC101 载体上去的 DNA 限制片段。随后，把这两种经过同一种限制酶切割消化的外源 DNA 和载体 DNA 混合起来，并加入 DNA 连接酶，由于它们具有同样的 EcoRI 黏性末端，因此便能够通过碱基互补配，形成双链结合体。其中单链缺口经 DNA 连接酶封闭之后，便产生出稳定的杂种 DNA 分子（图 9－7）。

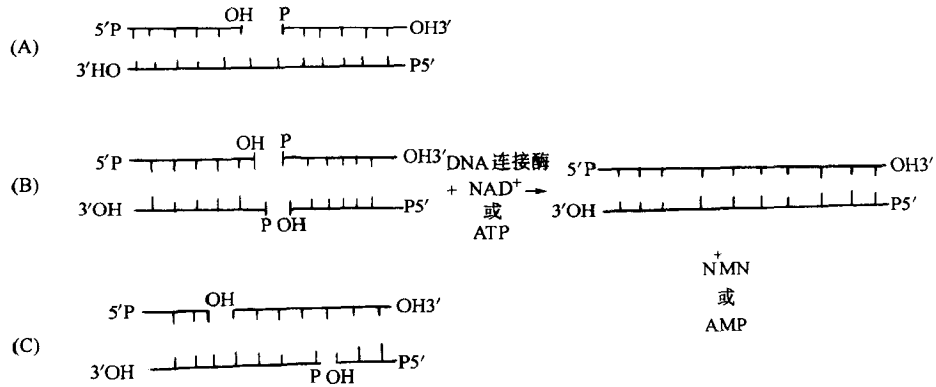


图 9－7　DNA 连接酶对缺口 DNA（A）、平末端 DNA 和黏性末端 DNA 的连接作用（B 和 C）

目前有三种常用方法可以在体外将供体 DNA 与载体 DNA 分子连接在一

起，产生重组 DNA 分子。第一种方法是用 DNA 连接酶连接具有互补性黏性末端 DNA 片段；第二种方法是用 T_4DNA 连接酶直接将平末端的 DNA 片段连接起来，或用末端脱氧核苷酸转移酶给具平末端的 DNA 片段加上多聚（dA）多聚（dT）尾巴之后，再用 DNA 连接酶将它们连接起来；第三种方法是先在 DNA 片段末端加上化学合成的衔接物，使之形成粘性末端之后，再用 DNA 连接酶将它们连接起来。这三种方法虽然各有差异，但共同的一点都是利用 DNA 连接酶所具有的连接和封闭单链 DNA 的功能。

同限制性内切核酸酶一样，DNA 连接酶的发现与应用，对于重组 DNA 技术学的创立与发展也具有头等重要的意义。它们都是在体外构建重组 DNA 分子所必不可少的基本工具酶。

（二）基因文库

虽然通过琼脂糖凝胶电泳可以从病毒染色体、细菌质粒或细菌染色体上分离到限制性片段，但采用此法并不能从高等真核生物的染色体上分离到单一限制性片段。因为一种限制酶可以将真核细胞的染色体 DNA 切割成许多条限制性片段，在凝胶电泳时，各种大小相近的片段相互混杂，而不能形成清晰的带纹。例如，限制酶 EcoR I 可以将长约 8×10^9bp 的人体基因组切割成约 20 万条大小不同的片段，片段数目如此之多，在凝胶上根本不可能加以分辨。

如果采用 DNA 重组技术，将所有染色体片段插入到克隆载体上，那么各种大小不同的片段，甚至长度非常相近的片段都可以分离出来。例如，将果蝇基因组的 4 万条不同的限制性片段分别插入到细菌质粒上，使每一个质粒只载有一条限制性片段，当转化大肠杆菌后，每一细菌菌落就只含有一条外源 DNA 片段。这种含有所有果蝇基因组片段的菌落群体就叫做基因文库（gene library）。在基因克隆过程中，最为困难的是如何从一个庞大的基因文库 中鉴别出目标 DNA 片段。现在，已有很多方法可用来选择目的基因。

利用 cDNA 为探针，与细菌菌落杂交是鉴定和分离目标克隆的重要方法之一。所谓 cDNA 即利用 mRNA 为模板和逆转录病毒的逆转录酶而复制的一种互补 DNA（complementary DNA）。逆转录过程需要引物，即一小段能与 mRNA3′端多聚 A 尾巴互补的寡聚 dT 序列。在存在四种脱氧核苷三磷酸的条件下，逆转录酶以寡聚 dT 为引物按 5′→3′方向合成一条单链互补 DNA。这种单链 DNA 往往能自发形成一种发夹环结构，逆转录酶又能以这种发夹环结构为引物，合成另一条互补 DNA 链。发夹环可以用 S1 核酸酶切除掉，然后再接上具限制酶切点的接头以创造粘性末端。当 cDNA 插入到克隆载体上后，再将这种重组 DNA 引入大肠杆菌，进行扩增（图 9－8）。

cDNA 含有编码蛋白质的遗传信息，但这种信息不完整，因为通常很难得

到完整的 mRNA。即使有完整的 mRNA，cDNA 也不可能完全的与染色体基因相同，它缺少某些控制序列如启动子、内含子或间隔序列等。为了研究基因的真实结构和调节表达情况，还必须获得染色体基因，所以 cDNA 是获得染色体基因所必需的探针。

（三）目标克隆的寻找与鉴定

1. 利用探针寻找目标克隆　基因文库中包含有上千万个 DNA 片段的克隆，如何发现我们所感兴趣的基因是一个关键的问题。

cDNA 是根据某种特化细胞中总体 mRNA 通过逆转录合成的，所以还需要有适当方法将目标 cDNA 克隆从 cDNA 文库中分离出来。筛选目标 cDNA 克隆的方法很多，也很复杂。利用化学合成的寡核苷酸作探针，筛选目标 cDNA 克隆是常采用的方法之一。采用此法，首先需要了解目标基因编码的蛋白质的氨基酸序列，然后再根据氨基酸序列及其相应的遗传密码，合成一段寡核苷酸序列。因此，这种化学合成的寡聚核苷酸就可以用作探针来筛选目标 cDNA 克隆。

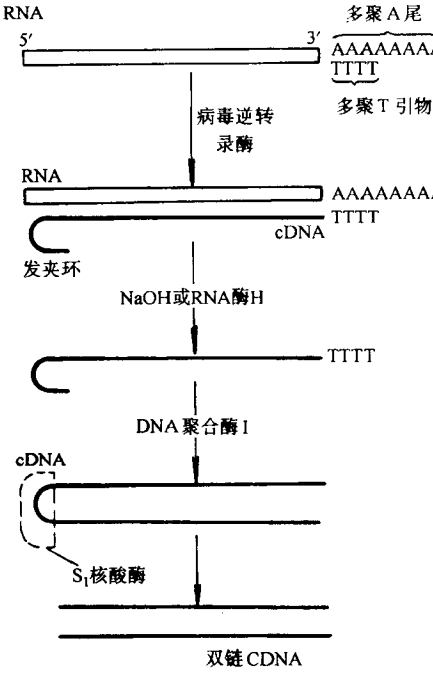


图 9－8　由 mRNA 通过逆转录形成双链 cDNA

一条短的 dT 链与 mRNA 的尾部的多聚 A 结合，作为逆转录的引物，以 mRNA 序列为模板逆转录合成 DNA；cDNA 在末端形成发夹环，而 mRNA 链可以用 NaOH 处理而降解；随后发卡环的一端作为引物，聚合酶 I 以新合成的 DNA 链为模板合成另一条 DNA 链；发卡环可用 S1 核酸酶降解

当获得适当的探针以后，采用菌落杂交法（colony hybridiztion）就能从 cDNA 文库或基因文库中把目标菌落筛选出来（图 9－9）。细菌菌落可以直接在硝酸纤维膜上培养，也可以从培养基转移到硝酸纤维膜上，但在实验过程中须保留母板或一个复制板。然后，将载有各种菌落的滤膜用 NaOH 处理，以裂解细胞和使 DNA 变性。将滤膜烘干以后，菌落中的 DNA 就被束缚到滤膜上，因此滤膜上每个菌落的位置就与原始菌落的位置相同。

在进行菌落杂交之前，还需要用 P^{32}标记探针。目前标记方法很多，但通常采用切口平移（nick translation）标记法和寡聚核苷酸标记法。标记后的 DNA 探针具有放射性，通过高温变性后同滤膜一起置适当缓冲液内培养，探

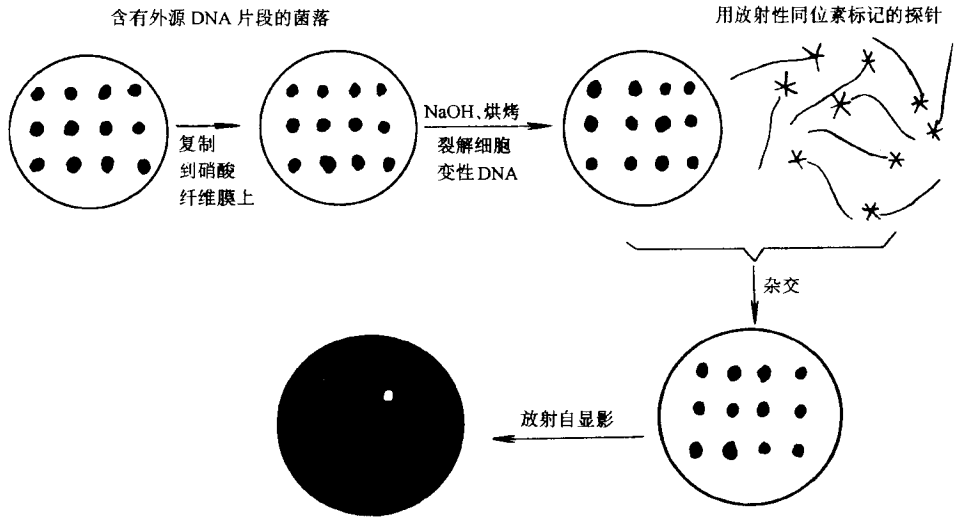


图 9-9　菌落杂交的基本过程

针就能与具有同源性的 DNA 序列互补配对。洗脱未结合上的探针以后，通过放射自显影，就可以见到探针结合的位置，再参考母板上各个菌落的位置，就可以分离到任何与探针有同源性的菌落。如果某个菌落与探针存在同源性，该菌落就含有目标 cDNA 或目标染色体基因。

采用菌落杂交法，一次就可以筛选几千个菌落，因此只要有适当的探针，就可以从基因文库中筛选到目的基因。采用上述方法分离、鉴定目的基因的过程叫鸟枪射击法（shotgun method），即把基因组中所有基因都包括在射击范围内，从中筛选出目的基因。

DNA 探针有数个来源，一个是从能表达目的基因的组织中获得 cDNA，然后以 cDNA 为探针筛选含目标基因的菌落。例如哺乳动物的网织红细胞中 90% 的 mRNA 是 β-珠蛋白基因，因此很容易从 mRNA 得到 cDNA 探针，运用该探针可以进一步探测基因组 DNA。如果所要了解的问题是 cDNA 所能提供的，那么 cDNA 克隆就足够了。但是如果需要该基因的内含子和调控序列，就必须从基因组 DNA 克隆整个基因。另一个方法是运用已知的、与克隆基因相似的、源于其他组织或动物的基因来设计探针。例如根据子囊菌中链包霉属（*Neurospora*）中的某一被克隆的基因序列来设计用于检测 *Podospora* 的同源基因的探针。这个方法依赖 DNA 序列在进化上的保守性。尽管探针 DNA 与所要检测的 DNA 不是完全一致，但是它们的相似性足以引起探针与 DNA 杂交。这种方法被戏称为“电话克隆”（“clone by phone”），因为如果你打电话

给你的同事，而他恰好从其他组织克隆了与你所要克隆的基因相似的基因，那么你的工作将变的非常容易了。还有一种方法是如果所要克隆的基因的蛋白产物的氨基酸序列是已知的，那么可以根据遗传密码来设计该基因的探针序列。由于大多数编码同一个氨基酸的密码子都不止一个，那么就会存在几种可能的DNA序列。为了解决这个问题要选择短且变化最少的序列作为合成DNA探针的信息。DNA的化学合成过程是一个渐进的过程，因此可以同时合成编码同一氨基酸序列的不同的核苷酸序列，在溶液中形成各种可能的核苷酸序列的混合物。如图9－10的核苷酸序列中有5个位点可以发生变化，所要合成的探针的核苷酸序列就有 $2\times3\times2\times2\times2=48$ 种之多。以这种混合物作为探针，其中正确的序列将会发现所感兴趣的基因。

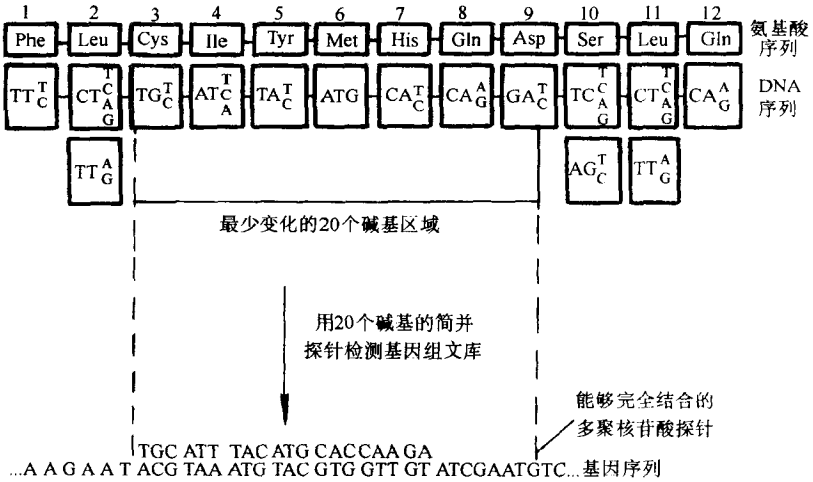


图9－10 目标蛋白质中的一段用于设计核酸探针的短的氨基酸序列

如果一个基因的产物蛋白质已被发现并被纯化出来，那么就可以通过制备寡核苷酸探针，检测基因文库中相应的基因，并且可以通过该蛋白的特异性抗体来检测某一组织的表达文库（expression library）。该文库是通过表达载体高水平表达一种细菌的蛋白质为基础建立起来的，cDNA插入该载体的细菌蛋白质表达框内，细菌将会产生融合蛋白。通过转膜将不同的克隆转到膜上，再将干燥的膜放入含有抗体的溶液中，产生该蛋白的细菌将会吸附抗体，然后再用标记了放射性同位素以及标记荧光素或显色反应的二抗，通过与标记方法相应的不同的手段来说明克隆中是否有相应的蛋白质基因的插入（图9－11）。

2．*通过功能性互补发现专一性克隆* 重组DNA分子的克隆还能通过恢复被转化的受体细胞的失去的功能而被检测出来。运用这种功能性互补（functional complementation）所进行的DNA克隆技术称为互补作用克隆技术。该技

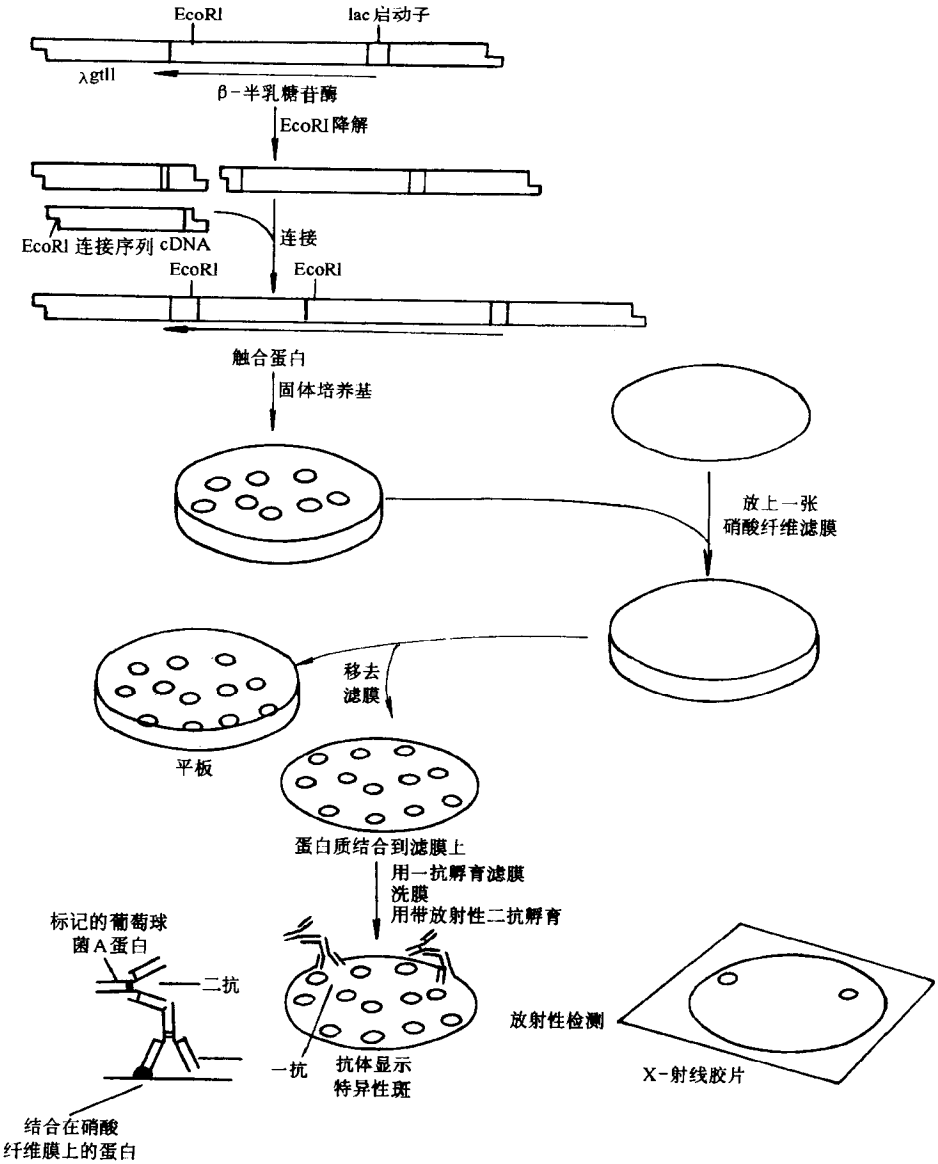


图 9-11　通过抗体鉴定目的克隆

图中的表达文库是用衍生于 λgt11 的噬菌体建立的，用蛋白质将异性抗体对其进行筛选。从滤膜上洗脱未结合的抗体后，就可通过与具放射性的第二抗体的结合情况观察已结合的初级抗体

术可以认为是经典微生物学方法的一种直接的扩展与延伸。如果被克隆的DNA片段同重组DNA分子是同源的，那么使用互补作用进行基因克隆就是一

种十分有效的方法。该方法是将大肠杆菌 DNA 的基因文库即一组含有大肠杆菌基因组全部 DNA 序列结构的重组 DNA 分子的异源群体导入一个营养突变株，随后将该突变株铺于缺少该菌株所需要的底物的基础培养基上。如果重组质粒 DNA 分子含有野生型的等位基因，营养突变型菌株就成为可在该培养基上生长的原养型菌株，而不含有该等位基因的菌株则不能在该培养基上生长，即只有那些获得了营养缺陷型互补基因的受体细胞才会长成转化菌落。

3. 定位克隆　定位克隆（positional cloning）是指利用基因的位置信息，对该基因进行有效克隆的任何方法，因为知道了基因的位置就可以避免通过分析整个基因文库来发现有用基因的繁琐过程。采用此法首先必须获得与靶基因紧密连锁的另一已克隆到的基因或标记，并以此作为确定靶基因的起点。然后以其末端序列作为从基因文库中选择与其相邻的克隆的探针。这样一步一步的进行下去，直至获得整个靶基因片段。这一程序叫做染色体步查法（chromosome walking），如图 9－12。

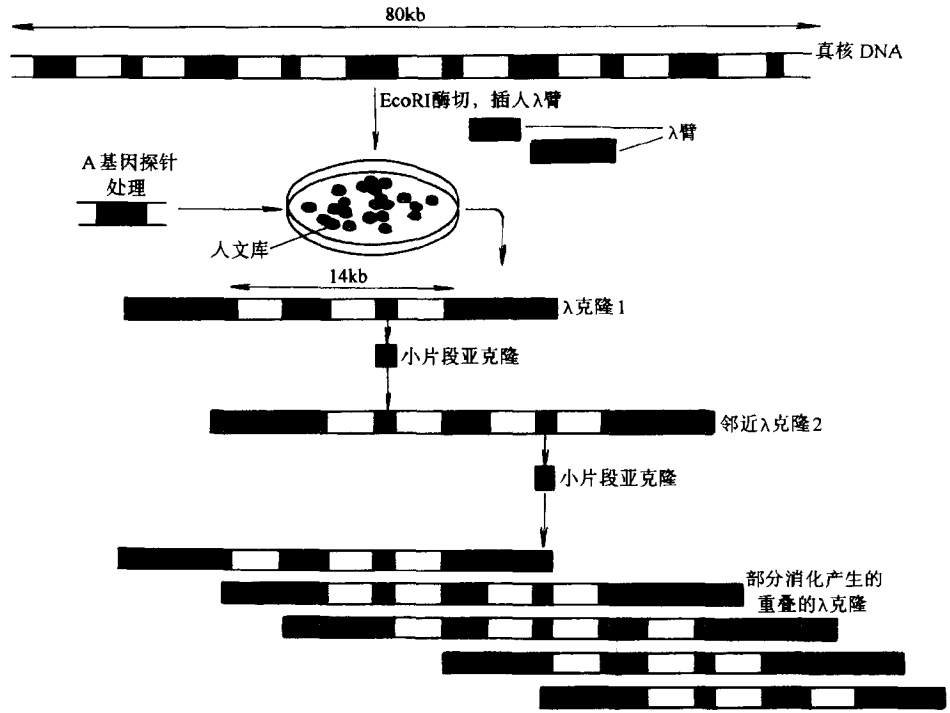


图 9－12　染色步查法

一个含有 EcoRI 降解的真核基因组 DNA 片段的噬菌体，可以用于分离另一个含有与该真核基因组片段相邻的 DNA 片段的噬菌体

4. 通过标签法克隆基因　标签法（Tagging）是通过将序列插入特定的靶基因而诱导靶基因突变，然后再以该序列作为标签或探针来检测目的基因。外源 DNA 通过转化或其他方法如显微注射可以通过重组而整合到基因组中，成为染色体的一部分。当外源 DNA 片段插入到某基因中或其附近时，将会引起基因突变，破坏该基因的功能。

假定用某种外源 DNA 片段转化细胞，转化体中就会有一部分为突变型。然后将转化体与正常个体杂交以确定突变表型是否随插入片段一起分离。如果是，就可说明该突变型是由外源插入片段引起的，该突变体的 DNA 就可用来建立基因文库，外源 DNA 片段就可用作探针来确定靶基因。为了获得完整的野生型基因，还必须将插入突变基因的亚克隆片段用作探针，从野生型基因文库中获得靶基因（图 9－13）。

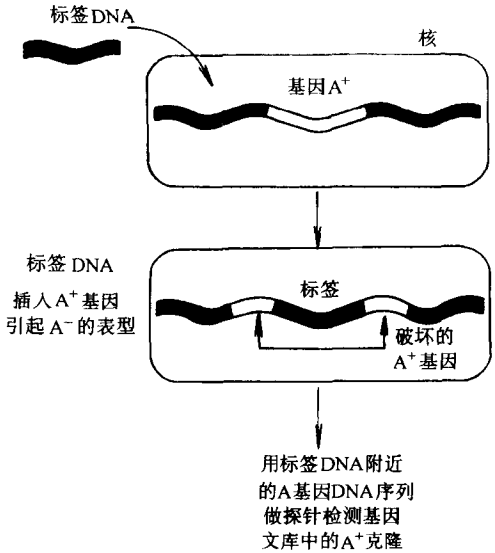


图 9－13　利用 DNA 插入片段作为标签搜寻目的基因

五、DNA 序列分析

DNA 序列分析法（DNA sequensing）是在核酸的酶学和生物化学的基础上创立并发展起来的一门崭新技术，对于从分子水平上研究基因的结构与功能的关系以及在克隆 DNA 片段的操作方面，都有着十分重要的实用价值。只有在详尽地掌握了 DNA 的核苷酸序列结构之后，我们才能够应用电子计算机程

序来标定出有关基因所包含的全部限制性内切核酸酶的识别位点，并据此构建出精确完整的限制图；也只有在详尽地了解到 DNA 的核苷酸序列结构之后，我们才能够用体外定点突变方法研究基因的结构与功能之间的关系。仅此数例就可以看出，DNA 的核苷酸序列分析技术确实已成为 DNA 分子克隆研究中最重要的方法之一。

1965 年，美国 Cornell 大学以 S.W.Holley 为首的科学家小组第一次完成了长度为 75 个核苷酸的酵母丙氨酸 tRNA 的全序列测定。其办法是利用各种 RNA 酶把 tRNA 降解成寡核苷酸，经分离纯化之后，再分别测定这些寡核苷酸短片段的核苷酸顺序。早期发展出来的 DNA 序列分析法都与在 RNA 序列分析中所用的方法有相似之处，但这些方法对于 DNA 序列测定并不适用。后来在 DNA 序列测定技术上出现的突破则是在独创性研究思想指导下取得的。1975 年，A.Maxam 和 W.Gilbert 发明了化学修饰序列测定法，F.Sanger 及其同事发明了双脱氧序列分析法。这两种方法都能准确分析一定长度的 DNA 序列。这里我们将以使用较广泛的 Sanger 法为基础，对 DNA 序列分析作扼要介绍。

双脱氧序列分析法（dideoxy sequencing melhod）又叫末端终止法，是由英国科学家 F.Sanger 于 1977 年发明的。这是一种简单快速的 DNA 序列分析法。他将单链 DNA 模版分到 4 个试管中，在每个试管中分别加入引物、DNA 聚合酶、四种脱氧单核苷酸和一种双脱氧单核苷酸（ddATP、ddTTP、ddGTP 和 ddCTP)。如果在试管中加入 ddATP 及其他四种脱氧核苷三磷酸，其中一种是带有^{32}P 放射性标记的，在 DNA 合成过程中一旦 ddATP 代替 dATP 掺入 DNA 分子，就可终止 DNA 链的合成。ddATP 在不同的位置加入，可形成长短不一的一系列 DNA，它们都具有同样的 5′末端，而且其 3′末端终止位点的核苷酸都应该是 ddATP。将这种混合物加到变性凝胶上进行电泳分离，就可以获得一系列全部以 3′末端 ddATP 为终止残基的 DNA 片段的电泳谱带。同样，使用相应于其他核苷酸的抑制物如在加入 ddTTP 的试管中也合成一系列长短不一、以 ddTTP 结尾的 DNA 分子（图 9－14)。4 个反应分别在不同的试管中温育，模板链是一条单链 DNA，通过将一段待测 DNA 片段克隆到 M13 噬菌体或其衍生载体上而形成。DNA 的合成由一个通用引物开始，该引物与 M13 的多克隆位点区相邻的区域互补。因此，无论被克隆的 DNA 片段序列如何，都可以使用相同的引物来进行 DNA 序列分析。4 个样品在凝胶上分 4 行电泳之后，就可观察到其相对长度，相应碱基的位置沿 4 个样品带阅读就可被确定下来（图 9－14)。现在，双链 DNA 分子也可以用双脱氧法直接进行序列分析。

迄今为止，已经报道过的采用引物合成进行DNA序列测定的各种方法都

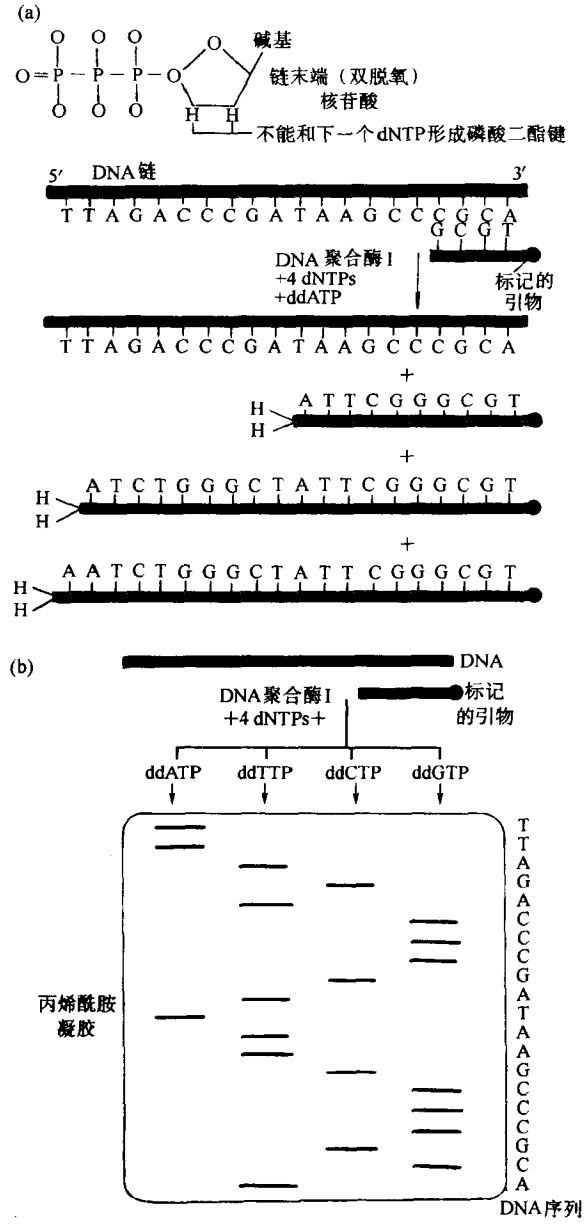


图 9－14　双脱氧法 DNA 序列分析原理示意图

A. 一段标记的引物（根据载体的侧翼序列设计）用于启始 DNA 合成，合成过程因随机掺入 4 种不同双脱氧核苷酸（此处为 ddATP）而被中止　B. 通过凝胶电泳将合成的片段分开，并进行放射自显影

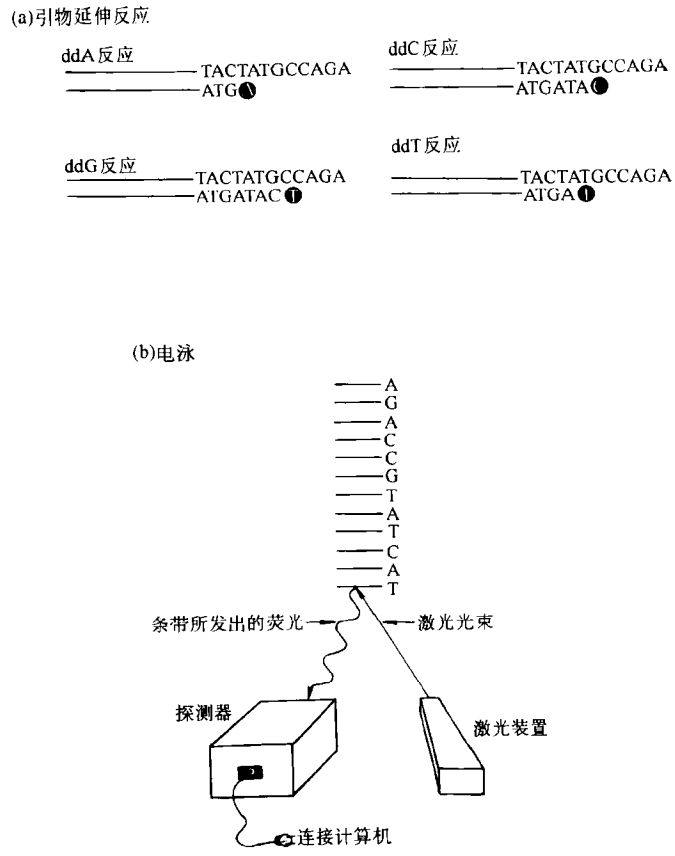


图 9－15　自动化的双脱氧法 DNA 序列分析

A. 引物延长反应的产物按手工法进行凝胶电泳，4 种双脱氧核苷酸分别用不同荧光分子标记，使之发出不同颜色的光　B. 电泳和确定带纹。各种引物延长反应的产物通过凝胶电泳被分开，每一条带依据荧光标记发出一定颜色的光，如黑色为 ddA、蓝色为 ddC 等。激光探测器分析荧光颜色。计算机将信息转换成碱基序列并贮存下来　C. 自动 DNA 序列分析实验的打印结果。不同颜色的峰代表每一条带的荧光强度

有一些共同特点，即需要将高分子量的 DNA 分子通过限制酶消化，降解成一组具一定长度的限制性片段，然后再用琼脂糖凝胶或聚丙烯酰胺凝胶电泳分离纯化，并最后从凝胶中抽提出 DNA 片段供进一步序列分析使用。然而，为了将这些降解的 DNA 片段按其原来的顺序“拼接”起来，至少还需要用另一种限制酶对同种大分子量的 DNA 作酶切消化，并进行电泳纯化和限制性片段作图。这些供作序列测定的 DNA 片段的长度，绝大多数都仅为数百个核苷酸。因此需要用物理方法分离大量的 DNA 片段。这显然是一种十分繁重的负担。这些工作无疑既花费时间又花费大量的金钱。为了保证能够获得所有的限制片段，就需要制备足够量的 DNA，而其中有许多步骤都要用不同浓度的凝胶进行电泳再纯化。在这种再纯化过程中，会加剧 DNA 的丢失和损伤。克服这些缺点的途径之一是采用 DNA 分子克隆的方法，即将不同限制性内切酶消化的 DNA 限制片段随机克隆到一种合适的载体分子上，以保证所有模板链都来自同一个重组体克隆。现在一般将待测的 DNA 片段克隆在 M13 载体上，获得理想的单链 DNA 模板。

但是，在大量的 DNA 片段的序列测定中，例如人类基因组工程中的 30 亿碱基对的序列测定运用上述的“人工”的方法则要花费很长时间，必须采用快速的自动化序列测定方法。图 9－15 示快速、自动测序方法之一，该法是根据 Sanger 法的原理建立起来的。该程序也同人工方法一样采用 4 种双脱氧单核苷酸（ddA，ddT，ddG 和 ddC），不同的只是 4 种双脱氧核苷酸分别加上不同的荧光素，当接受光照时就会发出不同颜色的荧光。然后我们进行 4 个反应，随后将 4 个反应液混合在凝胶的同一泳道上进行电泳。在凝胶背部安装一个可以用来发射激光束的分析仪，那么短核苷酸片段所发出的荧光将被分析仪所检测到并被输入计算机中，对所得到的结果进行分析。假如发出的荧光的颜色为兰色，这就意味着该短核苷酸序列的末尾碱基是 ddC，该碱基则为 C；同样，绿色意味着末尾的碱基是 A，菊黄色为 G，红色为 T。按顺序把核苷酸进行排列，就得出了该序列的核苷酸碱基组成（图 9－15）。同时，计算机可以打印出所记录的每一条荧光带纹和颜色以及它们所代表的相应的碱基组成。该方法既经济又简便、省力。

六、聚合酶链式反应

聚合酶链式反应（polymerase chain reaction，PCR）是一种从某种 DNA 混合物中有选择地复制某种目标 DNA 序列的方法。采用 PCR 技术首先需要含有被扩增的 DNA 序列的 DNA 混合物和一对寡核苷酸引物。这种引物接有一段

被扩增的 DNA 序列，它是人工合成的。因此，事先必须知道被扩增 DNA 的一部分序列。PCR 的程序如下：

1. 在 94℃下培养，将双链 DNA 变性成单链（图 9－16a）。

2. 成对地加入特定的引物并使之与单链 DNA 碱基配对（图 9－15b）。

3. 通过 DNA 聚合酶延长引物（图 9－16c）。这里所使用的 DNA 聚合酶为一种特殊的、叫做 Tag 的聚合酶，该酶从细菌（*Thermus aquaticus*）中分离而来，有很高的热稳定性，不会被 PCR 过程中反复的高温变性所破坏。*Thermus squaticus* 生活于热喷泉中，自然在这种细菌中就可分离到耐温性的酶类。

4. 按照要求重复上述三个步骤（图 9－16d－g）。

采用 PCR 技术，新合成的 DNA 量成几何级数增加。一个分子的 DNA 若经过 20 次循环，就可扩增到 $2^{20}=1048$，576 个拷贝。采用 PCR 仪只需几分钟就可完成一个扩增循环。

PCR 技术的一个主要缺点是 Tag 聚合酶不具有校对特性，所以扩增的 DNA 常有一定差错，但这种差错出现的频率很低。PCR 过程也容易受到污染，如果反应混合物中混有能与引物结合的其他 DNA，那么这种 DNA 也同目标 DNA 一样，一起被扩增。

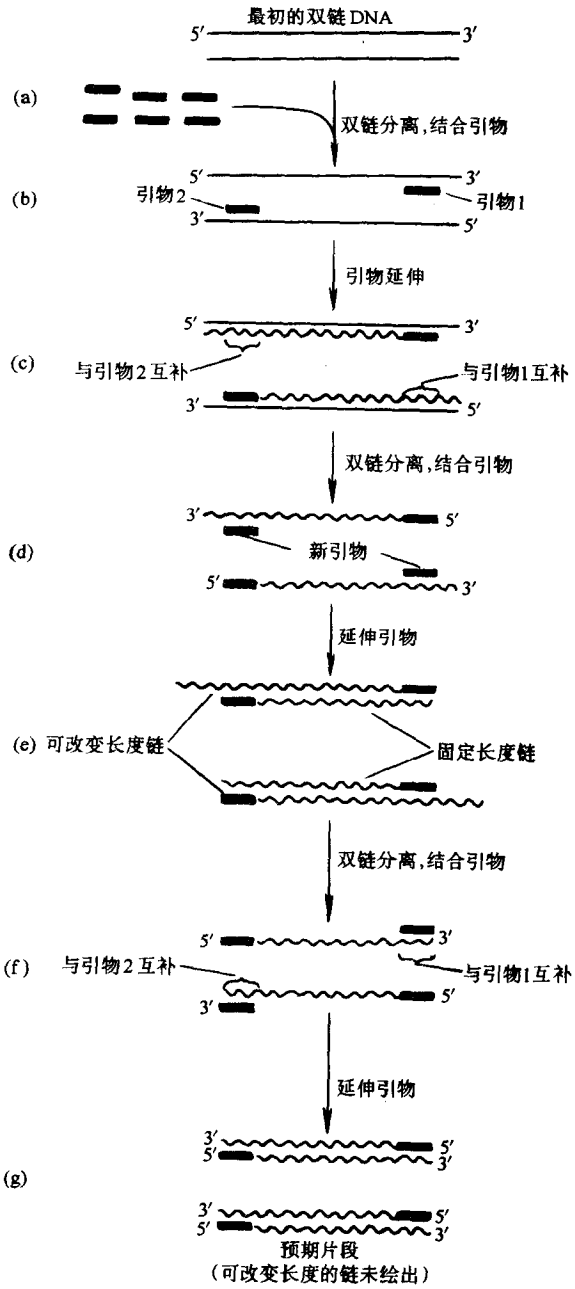


图 9－16 聚合酶链式反应的基本原理

PCR 技术应用很广泛，包括扩增克隆的 DNA、扩增用于序列分析、未克隆的基因组 DNA，以及扩增用于疾病诊断的 DNA 等。但无论哪种情况，事先都必须知道 DNA、序列组成的一部分信息，以合成适当的引物对。

第二节　基因操作的载体系统

作为载体，一个最基本的要求就是具有一个复制起始位点，能够在寄主细胞中进行自我复制，这样可使目的基因得到增殖，并传递到由这一寄主细胞所形成的后代中。目前常用的载体有质粒、噬菌体和黏粒等。下面将对这些载体作一些简要介绍。

一、质粒载体

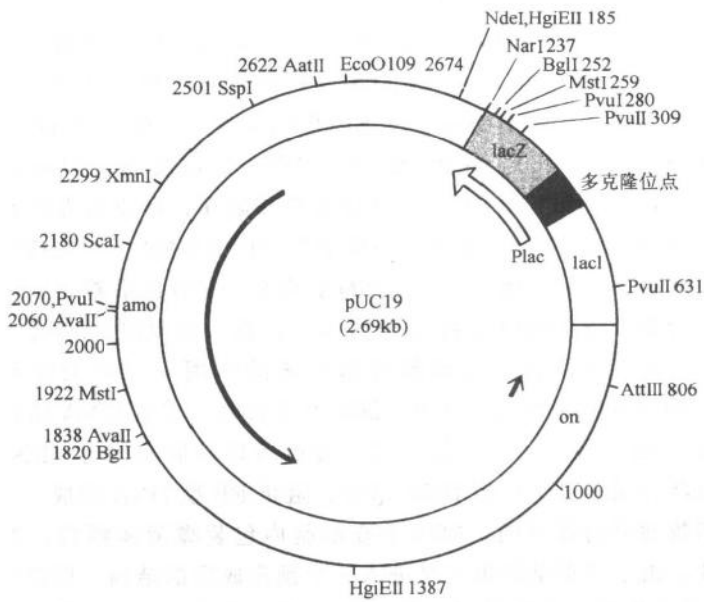
质粒是一种双链环状 DNA 分子，它带有一个复制起始位点，能够在细菌中独立进行自我复制。质粒在细菌中的复制方式有两种类型，一种称作松弛型（relaxed mode），一种称作严紧型（stringent mode）。目前作为载体的大部分质粒都是松弛型的，它们起源于质粒 pMB1。在正常生长情况下，带有这一复制原点的质粒在每一个细菌细胞中至少有 15～20 个拷贝。这种质粒复制的一个最大特点就是在复制时不需要其自身所编码的基因产物，而完全依靠寄主所提供的酶系统。因此，即使没有蛋白质形成，这些酶还可以起作用。在加入大量氯霉素等抗生素时，蛋白质的合成被抑制，从而阻止了细菌染色体的复制，而带有 pMB1 复制原点的质粒可以继续复制，在一个细胞中可增加到 2000～3000 个拷贝。相反，带有严紧型复制原点的质粒，如 pSC101 的复制需要一种自身所编码的蛋白与复制起始位点作用，而这种蛋白又负向调节它本身基因（repA gene）的转录，所以在一个细胞中这种质粒一般只有 1～5 个拷贝。正是由于这种质粒的复制需要由质粒所编码的蛋白质进行持续合成，所以抑制蛋白质合成的抗生素也能抑制质粒的复制。

分子克隆一般都是利用松弛型载体，因为这样在单位体积的培养物中可以得到较多的 DNA 和较多的蛋白质产物。但在某些情况下，比如蛋白质产物是有害的，含量过高会致死，我们就必须用“严紧型”克隆载体。

在自然条件下，许多质粒载体（plasmid vector）从一个细菌转移到新的寄主是通过细菌的接合（bacterial conjugation）实现的。这一过程需要一个 mob 基因（mob gene），而现在大多数作为载体的质粒都不含这一基因。在实验室操作中，质粒进入新的寄主是通过转化进行的。即使在最佳条件下，也只有少部分

细菌能够吸收外源 DNA，成为转化体（transformant）。为了区别转化和非转化的细菌，一般在质粒上需要有一个选择标记基因。常用的选择标记基因多为能够使细菌对抗生素如氨苄青霉素、四环素、氯霉素和卡那霉素等产生抗性的基因。

许多天然存在的质粒往往不能满足分子克隆对载体的要求，只有经过改造才能成为优良的载体。一般载体构建需要考虑以下因素：(1) 质粒载体的分子量应尽可能小。因为当质粒达到 15kb 以上时，转化率大大降低。另外小质粒还具有易于提取、易于鉴定和不易断裂等优点。(2) 质粒能够在宿主细胞中大量复制，以便得到大量的重组 DNA 分子。这也就是我们前面所提到的现在所用的质粒大多用松弛型复制子的原因。(3) 载体应该包含一个可供选择的标记基因，以便选择转化体。(4) 具有多克隆位点（multiple cloning site，MCS），就是在一个位置上有多个限制性内切酶的切点便于克隆工作。图 9－17 是一个在分子克隆中常用的质粒 pUC19 的结构。



多克隆位点：

1	2	3	4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	5	6	7	8
Thr	Met	He	Thr	Pro	Ser	Leu	His	Ala	Cys	Ars	Ser	Thr	Leu	Glu	Asp	Pro	Arg	Val	Pro	Ser	Ser	Asn	Ser	Leu	Ala
ATG	ACC	ATG	ATT	ACG	CCA	AGC	TTG	CAT	GCC	TGC	AGG	TCG	ACT	CTAGAG	GATCCC	CGG	GTACCG	AGC	TCG	AAT	TCA	CTG	GCC		

HindIII　SphI　AstI　SalI AccI MincII　XbaI　BamHI　SmaI XmaI　KpnI　SacI　EcoRI

图 9－17　pUC19 质粒的结构

上图示 pUC19 的物理结构图和有关限制性内切酶的切割位点及其相应的碱基位置。下图示多克隆位点的各种限制性内切酶的识别序列

二、单链噬菌体克隆载体

大肠杆菌的丝状噬菌体有M13、f1和fd，这三种噬菌体的染色体组是高度同源的（98%以上）。由于是单链DNA分子，作为克隆（cloning vector）载体具有以下可能的用途。如大量扩增用于序列分析的单链DNA，合成只有一条链被标记的DNA杂交探针等。

（一）M13噬菌体的生物学

如第6章所述，M13是一种丝状的大肠杆菌噬菌体，它包含一个单链的环状DNA分子，全长6407个核苷酸。基因组的90%编码对噬菌体生存所必需的蛋白产品，共有10个基因，大部分基因之间只有少数非编码核苷酸。但在基因Ⅱ和Ⅳ之间及基因Ⅷ之间有较长的基因间区。这些区域有控制基因表达和病毒DNA合成的调控序列。

M13噬菌体在自然条件下是通过性伞毛（sex pili）侵染的，而性伞毛的形成是由大肠杆菌中F因子（一种附加体）控制的。M13的单链DNA（+）进入大肠杆菌细胞后，在细胞内酶系统的作用下形成双链环状DNA分子，称作亲本复制型DNA（RF）。这种RF型的DNA可以复制，形成新的复制型DNA。复制的同时，（-）链的基因开始转录和翻译，形成病毒的各种mRNA和蛋白质。M13噬菌体基因Ⅱ的蛋白能够在RF型DNA（+）链的特定位置创造缺口，然后，大肠杆菌DNA聚合酶Ⅰ按5′→3′方向沿着（-）链加入互补核苷酸，合成一条新链以取代原来的（+）链。当复制成环时，被替代的（+）链切下来重新环化，在细胞内酶系统的作用下又形成闭环的RF型DNA。当细胞中的噬菌体达到100～200个拷贝时，单链DNA结合蛋白（基因Ⅴ的产品）已积累到一定的量，它一方面可以抑制基因Ⅱmrna的翻译，另一方面与新合成的（+）链DNA结合，阻止RF型DNA形成。

与大多数细菌病毒不同，M13不在细胞内包装噬菌体颗粒，而在细胞围膜间隙包装。由于病毒基因组不是进入一个预先成形的结构，所以可以包装的单链DNA的长度没有严格限制。在M13中发现有包含好几个拷贝的基因组的病毒粒子，也克隆过相当于好几个病毒基因组长度的外源DNA片段。

（二）M13噬菌体的改造

Messing等将含有大肠杆菌LacZ基因的控制序列和缺失的LacZ基因（编码β-gal酶氨基端146个氨基酸的一段DNA）插入基因Ⅱ和Ⅳ之间的非编码区，构建了M13mp系列载体（图9-18）。寄主中的F因子上也带有1个突变的Lac Z基因，它编码一个氨基端第11～41个氨基酸缺失的多肽。这两个缺

失基因的产物都没有活性。但若将两者混合，就能形成有活性的β－半乳糖苷酶。这一现象被称作α－互补（α－complementation）。当培养被M13mp1噬菌体侵染的大肠杆菌菌系时，如果在培养基中加入IPTG和X－gal，在β－半乳糖苷酶的作用下，就会形成兰色的噬菌斑。如果在M13mpl的多克隆位点又插入一个外源DNA片段，就破坏了α－供体的形成，细胞中不能产生有活性的β－半乳糖苷酶，故重组体M13mp1所形成的噬菌斑为白色或无色。利用这种颜色变化就很容易鉴别重组体和非重组体。

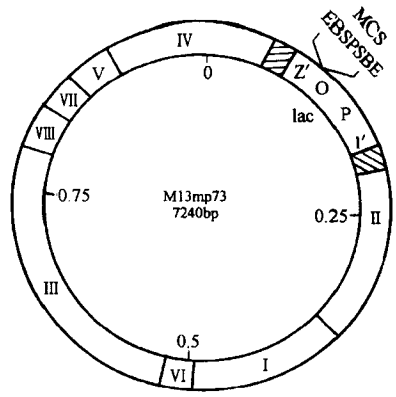


图9－18　改造的M13mp系列载体之一M13mp73的结构

现在所用的pUC系列质粒载体也利用了α－互补的原理，也可利用颜色变化识别重组体。

三、噬菌粒——含有丝状噬菌体复制起始位点的质粒

在M13中进行克隆往往存在两个问题：（1）外源DNA部分片段易丢失，片段越长越易丢失。（2）外源DNA往往只以一个方向插入。在这种情况下，用一个带有M13复制起始位点的质粒就可解决这一问题，这种重组DNA叫做噬菌粒（phagemid）。现已广泛使用的商标名称叫做pBluescript（pBS）的载体为例，对噬菌粒的结构特征进行说明（图9－19）。

同pUC载体系统相似，pBS噬菌粒也具有一个插入到lacZ基因中的多克隆位点，所以含有插入序列的克隆很容易通过由X－gal染成蓝、白两种颜色加以鉴别。这种载体还有一个单链DNA噬菌体f1（与M13相近）的复制起始区，也就是说含有噬菌粒的细胞可以产生和包裹单链噬菌粒DNA。这类载体的另一个重要特征是多克隆位点两侧有两个不同的噬菌体RNA聚合酶启动子。如pBS在该位点一侧有T3启动子，另一侧有T7启动子，以便分离双链DNA和在离体条件下利用其中任何一种噬菌体的RNA聚合酶合成高纯度的与其中一个单链互补的RNA转录子。

图9－19中，pBS源于质粒pBR322，具有其氨苄青霉素抗性基因（Amp^r）和复制起始区（ColE1 ori）。它还有噬菌体f1的复制起始区（f1 ori），所以如果有f1帮助噬菌体为其提供复制机器，pBS就能以单链形式分泌到细胞外。多克隆位点（MCS）在T7和T3两个启动子之间有21个单一限制性位

点，因此任何插入的 DNA 片段都可以在离体条件下进行转录，合成与两条单链之一互补的任何 RNA。MCS 位于大肠杆菌 lacZ′基因之中，在加入该诱导物如 IPTG 时可以消除 lacI 基因产物的抑制作用，使未被切割的载体在加入了指示剂 X－gal 时能合成 β－半乳糖苷酶的 N－末端片段，克隆呈蓝色。如果克隆含有在 MCS 中插入了 DNA 片段的重组质粒，lacZ′基因失活，不能合成β－半乳糖苷酶，从而使克隆呈白色。

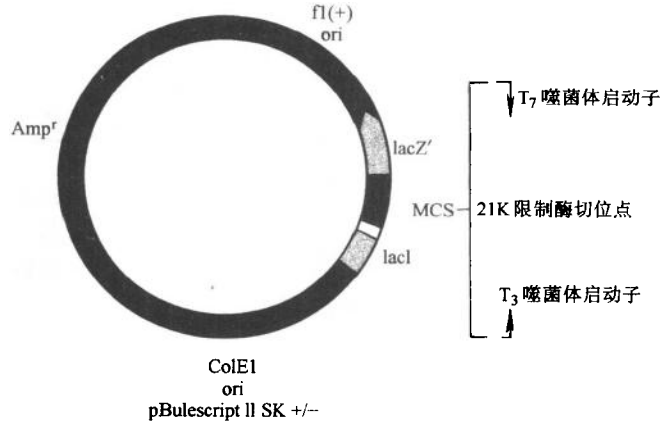


图 9－19　噬菌粒 pBluescript 的结构

噬菌粒作为载体具有以下特点：（1）像质粒的双链 DNA 那样比较稳定，易大量提取；（2）克服了从质粒到 M13 载体的烦杂的 DNA 片段克隆工作；（3）因为噬菌粒本身很小，它可以克隆较长的 DNA 片段（可达 10Kb）。

其他一些噬菌粒如经过改造构建的 pUC118 和 pUC119 也已被广泛用作 DNA 的克隆。它们分别来自于 pUC18 和 pUC19，并带有一个来自野生型 M13 的 476 bp 的 DNA 片段。噬菌粒 pUC118、pUC119 通常和 M13K07 配合使用。

四、双链噬菌体载体

λ 噬菌体的基因组是双链线状 DNA 分子，长约 49kb，在两端带有 12 个核苷酸的单链互补末端，又称黏性末端。λ 噬菌体进入寄主细胞后，在可进入裂解周期，又可进入溶源化状态。线状的 λDNA 进入寄主细胞后，两个互补的黏性末端相连，在寄主连接酶作用下，补平缺口，形成环状 DNA 分子。环状的 λ 噬菌体开始复制，产生大量双链 λDNA 分子。在裂解周期的后期，这些双链 DNA 分子被包装进入蛋白质的外壳，形成完整的 λ 噬菌体。寄主细胞裂解后，释放出大约 100 个成熟的噬菌体颗粒。

关于λ噬菌体侵染循环的详细机制见第八章。但这里须指出，λ噬菌体外壳蛋白在包装时对基因组的长度有严格的限制，即要求包装的DNA片段不能短于原基因组的75%（36.8kb），也不能长于原基因组的105%（51.5kb）。这些因素都是将λ噬菌体改造成载体时所必须考虑的一些问题。

λ噬菌体DNA在大肠杆菌中复制和包装不需要噬菌体基因组的中间部分，所以中间部分即所谓填充片段（stuffer fragment）可以用限制酶EcoRI将其切除掉。留下的两臂然后再与插入片段连接在一起。这种重组的DNA分子既可通过转染被引入大肠杆菌中，也可通过离体包装，形成具侵染能力的噬菌体而进入大肠杆菌中（图9－20）。尽管两臂也可能自发连接在一起而不接受插入片段，但由于其DNA太小，不能被包裹到噬菌体头部内，所以不能形成克隆。

为了使λ噬菌体能更好地适合作载体的需要，一般还要进行以下几方面的改造。(1) 生物学制约：就是将无义突变（如琥珀突变）引进λ噬菌体裂解周期所需的基因中，构成许多安全的λ载体。在这些重要基因的编码区引入UAG终止密码，就会产生截短的肽链，使基因产物失去原有的活性。因而带有这种突变的λ噬菌体不能在正常寄主中生存。这样就避免某些特殊的或有潜在危害的重组体在自然条件下传播。而这种带有无义突变的λ噬菌体可以在有校正基因的大肠杆菌菌株中生存。因为这些菌株中可以产生一些突变的tRNA，它们能够识别突变的UAG，加入适合的氨基酸，从而形成完整的蛋白。(2) 载体的通用性：早期的λ载体如卡隆载体（charon vector）只能插入EcoRⅠ酶切的外源DNA片段，而这对于实际克隆工作是远远不够的。因此，人们构建了具有不同内切酶位点的λ载体，使λ载体具有更广泛的克隆能力。例如λWES、λB可以克隆长1～15kb的片段。外源DNA与λ载体重组时可以是插入，也可以替代原有的填充序列，这样使载体克隆的外源DNA长度有了更大的选择。

直接用重组λ噬菌体DNA去转染寄主细胞效率较低，一般每μgλDNA可形成约10^3～10^4个噬菌斑。如果将重组λ噬菌体DNA包装到噬菌体的外壳中，每μg重组体分子就可形成10^6个噬菌斑。所以，通常在利用λ噬菌体作载体时是先在体外包装重组噬菌体颗粒，再侵染寄主细胞（图9－20）。

重组噬菌体的选择有许多途径。其中之一是在填充序列的位置插入β－半乳糖苷酶基因。外源基因插入后，破坏了β－半乳糖苷酶基因的活性，从而可以通过噬菌斑颜色的变化判断是否为重组体。

一些λ噬菌体载体不仅能够克隆外源基因，而且能够让这些外源基因在细菌细胞中表达。如常用的λgt11和λgt18－23，它们带有大肠杆菌β－半乳糖苷酶基因的调控序列及其部分编码序列。在这个基因的编码区有EcoRI切点，

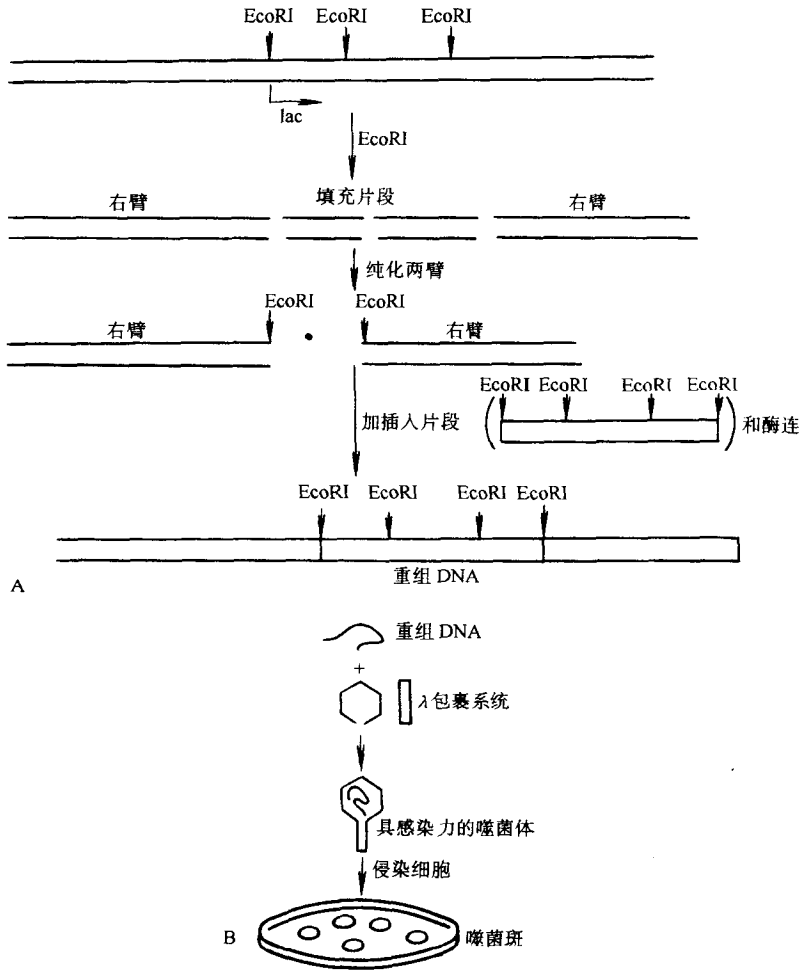


图 9－20　利用卡隆－4 载体克隆 DNA

A. 制备重组 DNA，用 EcoRI 切割载体，除去填充片段，保留 λ 的两条臂。然后将部分消化的插入 DNA 片段与 λ 的两条臂连接在一起　B. 包裹和克隆的重组 DNA 与 λ 包装元件混合，形成有功能的噬菌体粒子，最后将病毒粒子涂布到大肠杆菌上，收集噬菌斑

可以插入外源 DNA。在适合的寄主中，可以诱导嵌合基因表达，合成β－半乳糖苷酶的氨基端部分和由外源基因编码的多肽的融合蛋白（fusion protein）。

五、黏粒载体

黏粒（cosmid）是一种含有 λ 噬菌体 cos 区域的质粒（图 9－21）。这种载

体的设计是为了克隆和繁殖大片段的染色体 DNA。λ 噬菌体载体克隆外源 DNA 片段最长不超过 24kb，过长会造成染色体 DNA 包装困难。在实际工作中发现，高等真核生物的许多基因由于带有多个大型的内含子，片段很长，而不能以一个完整的基因在 λ 噬菌体或质粒载体中克隆和繁殖。比如鸡胶原蛋白基因、鼠二氢叶酸还原酶基因和人类固醇硫酸酯酶基因等的长度都达 40kb 以上。而粘粒作为载体则结合了质粒载体的优点和 λ 噬菌体克隆长片段的优势，使克隆片段长度可达 35～45kb。粘粒载体的这一特性在建立基因文库，通过染色体步查克隆目的基因等方面展示了广阔应用前景。

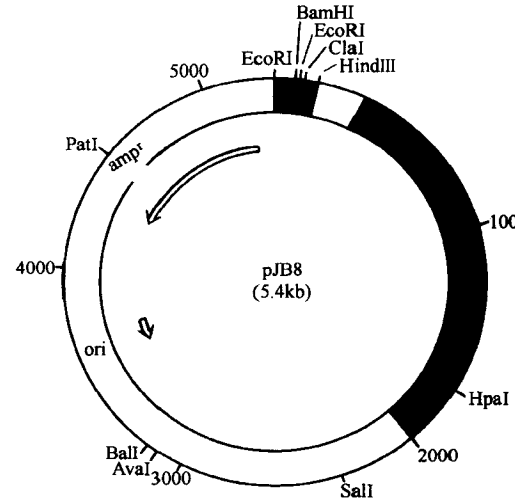


图 9－21　粘粒载体的结构

六、酵母人工染色体和细菌人工染色体

（一）酵母人工染色体（Yeast Artificial Chromosome）YAC

自克隆到酵母的着丝粒和端粒以后，1987 年 Burke 首次利用这两种结构建立了酵母人工染色体克隆载体系统，其所容纳的外源 DNA 比粘粒和噬菌体大 10 倍，可插入高达 1000kb 的 DNA 片段。

YAC 是由酵母染色体的两个与染色体完整性有关的端粒、一个与保证染色体正常分离的着丝粒、以及负责 DNA 与染色体复制的自主复制中心装配起来的人工染色体。如图 9－22，YAC 是由质粒 pBR322、酵母的着丝粒和一条染色体的左右两个端粒、酵母的自主复制序列以及一些选择标记基因构成的。该 YAC 在呈环状结构时，可以质粒的方式在原核细胞中增殖复制，酶切成线状以后，可连同插入的外源 DNA，如真正的染色体一样在酵母细胞中复制和被分配给到子细胞中。

以 YAC 作为载体，可以克隆平均长 300kb 的外源 DNA。人的基因组有 5～10万个基因，平均 30～50kb 有一个基因。因此，以 YAC 很可能克隆到包含一个完整基因的 DNA 片段，用任一基因旁或基因内的片段为探针便能筛选到含这一基因的 YAC 作载体的克隆。以定位克隆技术分离的与疾病相关的基

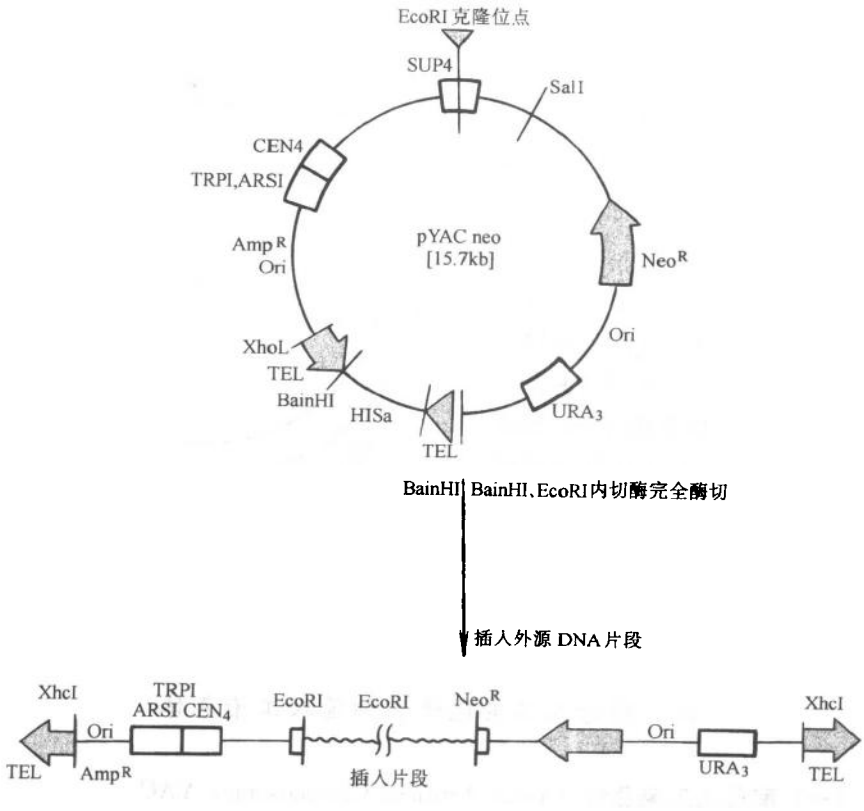


图 9－22　酵母人工染色体（pYACneo）的结构和克隆外源 DNA 示意图

TRP，URA，HIS：酵母基因；EcoRI，SalI，XholI，SamHI：限制性内切酶；TEL：端粒；CEN：着丝粒；Neo^R：新霉素抗性；Amp^R：氨苄青霉素抗性；ARS：自主复制序列；Ori：复制起点

因几乎无不从 YAC 克隆起步，其中包括近两年人类基因组研究的重大进展、亨氏基因和乳腺癌基因的克隆等。

构建 YAC 时，在原核生物的复制起点（Ori）上游连接一个限制酶切点，当 YAC 连有外源 DNA 片段后，只需用 XhoI 酶切，则靠近克隆位点的外源 DNA 上的 XhoI 切点可与 YAC 的 XhoI 切点互补形成环状，以质粒的方式进行复制。用这种方法可以把克隆的外源的 DNA 片段的末端分离出来，作为钓取与其邻接的、具有共同末端序列的即重叠序列的另一段 DNA 片段的探针。实际上这也是以染色体步行查法克隆基因。

（二）细菌人工染色体

细菌人工染色体（bacterial artificial chromosome，BAC）也是一种较新的

克隆大片段 DNA 的载体。BAC 源于大肠杆菌的 7kb 的 F 因子。由于 F 因子能携带大片段的大肠杆菌染色体片段，成为 F′因子，所以它同样能携带长达 300kb的外源 DNA 插入片段，但一般只插入 100kb 左右的片段。图 9－23 示较典型的 BAC 结构。虽然用 BAC 克隆的片段没有 YAC 大，但它仍有几个优点，如它能按细菌质粒的操作方法进行分离，而且很少产生杂合插入片段。

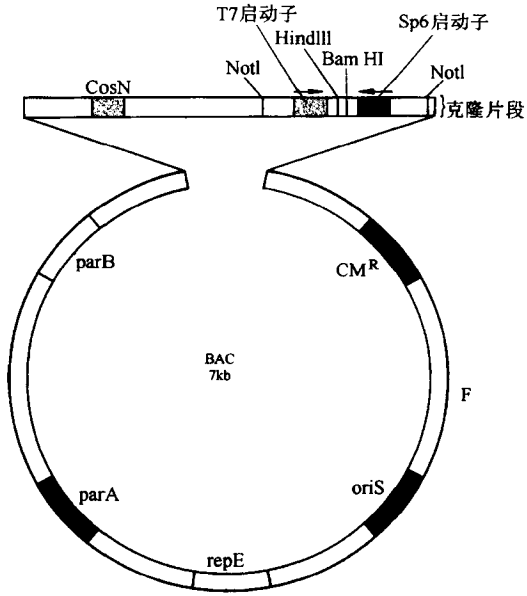


图 9－23　细菌人工染色体（BAC）的结构

CM^R 为氯霉素抗性选择标记，oriS、repE、parA 和 parB 为 F 因子的与复制和调节拷贝数有关的基因；CosN 为 λ 噬菌体的 cos 位点，Hind Ⅲ和 BamH Ⅰ为插入外源 DNA 的克隆位点；T_7 和 Sp6 两个启动子用于转录插入片段；Not Ⅰ位点用于将插入片段切割下来

七、表达载体

表达载体（expression vector）为一类能使克隆的基因进行转录和翻译的载体。例如 pUC 和 pBS 这两类载体都是将插入的 DNA 片段置于位于多克隆位点上游的 lac 启动子控制之下，如果插入片段位于 lacZ′基因可读框内就会阻断该基因，从而合成融合蛋白。融合蛋白的氨基末端的部分序列为 β－半乳糖苷酶的序列，羧基末端部分为由插入 DNA 片段编码的多肽序列（图 9－24）。

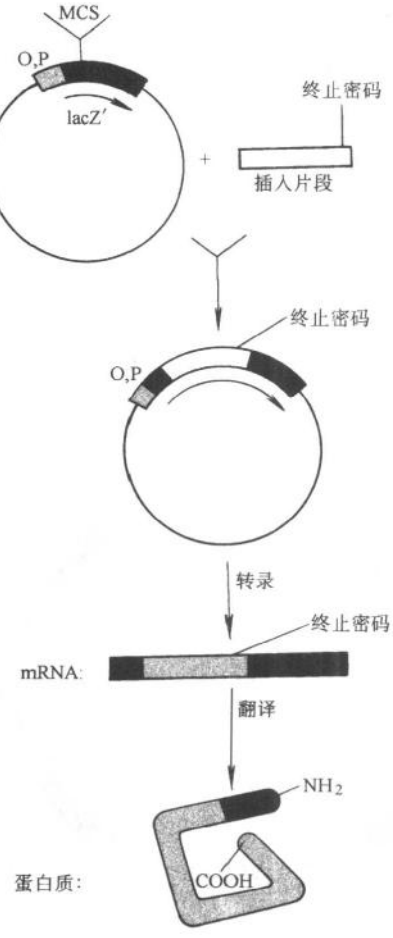


图 9－24　由 pUC 克隆的融合蛋白的形成

如果希望克隆的基因进行高水平表达，则需要使用特殊表达载体。这类载体通常都有一个强启动子和在起点密码 ATG 附近有一段 SD 序列。trp 操纵元的启动子就是细菌的强启动子之一，几种不同表达载体都使用了这一启动子如 ptrpL1（图 9－25）。

真核生物的蛋白质在细菌中大量合成往往对细菌本身是有毒的，即使这些蛋白质无毒，因大量累积也会干扰细菌生长。无论是哪种情况，如果克隆的基因始终表达，都不能使载有外源基因的细菌很好地生长，不能达到预期目标。如果将克隆的基因置于一诱导型启动子控制之下，则可控制外源基因的表达。

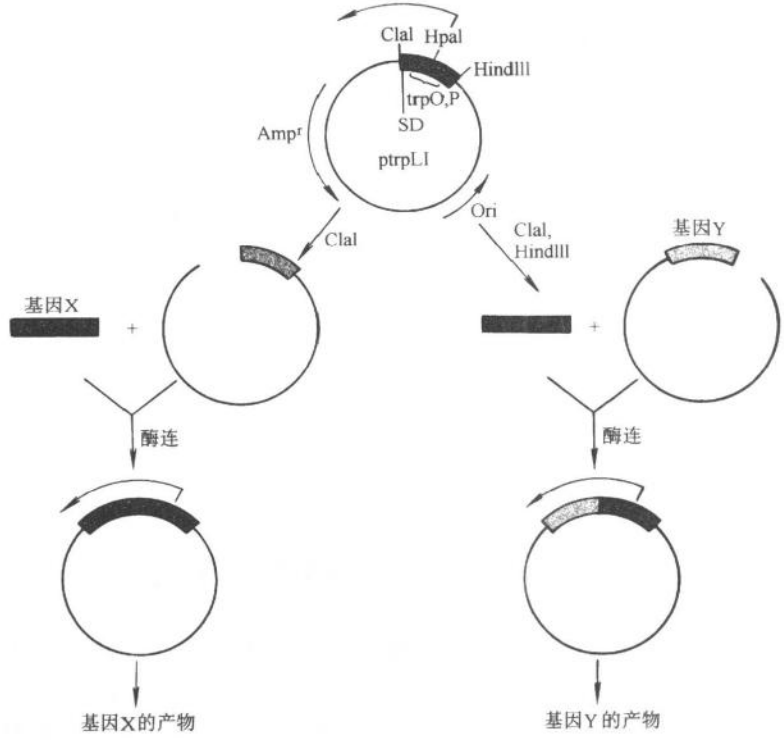


图 9－25　ptrpL1 表达载体的特征

这一载体在 SD 序列及操作子/启动子（trpO.P）序列之前有一个 ClaⅠ克隆位点，转录按逆时针（箭头）方向进行；在 ClaⅠ位点插入外源基因（X），该载体就可成为表达载体；另一方面，trp 的控制序列可以用 ClaⅠ和 HindⅢ酶切插入到另一质粒的编码序列（Y）之前

Lac 启动子可以被异乳糖或它的类似物 IPTG 所诱导，但是 Lac 抑制物的抑制作用是不完全的，即使在没有诱导物存在的情况下，克隆基因也可以表达。克服该问题的方法之一是将克隆的基因插入自身含有 lacⅠ基因的质粒或噬菌粒中如 pBS 质粒。如图 9－26，这个由载体产生的额外的抑制物可以使克隆基因处于关闭状态，直到用 IPTG 来诱导，克隆基因才能表达。

另一个策略是使用严密控制的启动子如 λ 噬菌体启动子 P_L。用该启动子的表达载体系统转入带有温度敏感 λ 抑制基因（cl857）的宿主细胞，如果我们保持这些细胞于 32℃ 以下，抑制基因发生作用，基因不进行表达。当我们升高温度到 42℃ 时，温度敏感抑制基因不再行使功能了，克隆基因开始表达。图 9－26 说明了表达载体 pKC30 的作用机理。

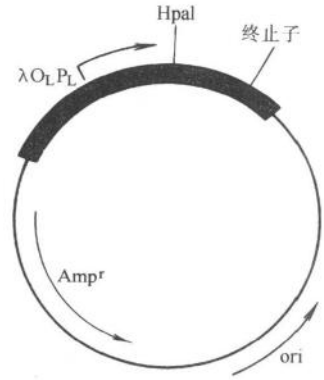


图 9－26　诱导型表达载体 pKC30 的结构

基因被插入到 O_LP_L 操作子/启动子下游的 Hpa Ⅰ 位点上。宿主细胞为载有温度敏感型 λ 抑制基因（cI857）的 λ 溶原性细胞，当温度从 32℃ 提高到 42℃ 时，温度敏感型 λ 抑制子失活并从 O_L 上脱离下来，基因开始转录

第三节　酵母的遗传工程

酵母（*Saccharomyces cerevisiae*）被看作是一种真核生物的大肠杆菌，其中一个主要原因是在经典遗传研究中酵母遗传行为得到了很好的研究和发展，现已明确了数千种突变型和由其决定的几百种不同的表型，这些成就为展开其分子水平的研究提供了丰富的遗传标记。酵母研究中的另一重要成就是发现了环状的 6.3kb 的酵母质粒即 2μ 质粒，该质粒可以在酵母的减数分裂和有丝分裂中进行传递，它是现今构建的许多酵母载体的基础材料。

一、酵母载体

最简单的酵母载体源于细菌质粒（图 9－27a）。这类质粒转入酵母细胞中后，可以与染色体发生同源重组而插入到染色体中。通过单交换或双交换，或者是在染色体上插入完整质粒，或者是将质粒上的等位基因替换染色体上的靶基因（图 9－28）。这些重组体均可以通过将细胞培养在对质粒上的基因具有选择作用的培养基上而被检测出来。这是因为细菌质粒不能在酵母中复制，因此只有同源重组才是惟一的可以产生稳定的转基因表型的途径。

以 2μ 质粒为基础将细菌质粒片段和酵母部分 DNA 片段拼接起来（图 9－27b）的结构有几大优点。首先，2μ 质粒的部分序列可以赋予新建质粒在酵母中进行自主复制的能力。其次，目的基因既可以被转入酵母中进行研究，也可

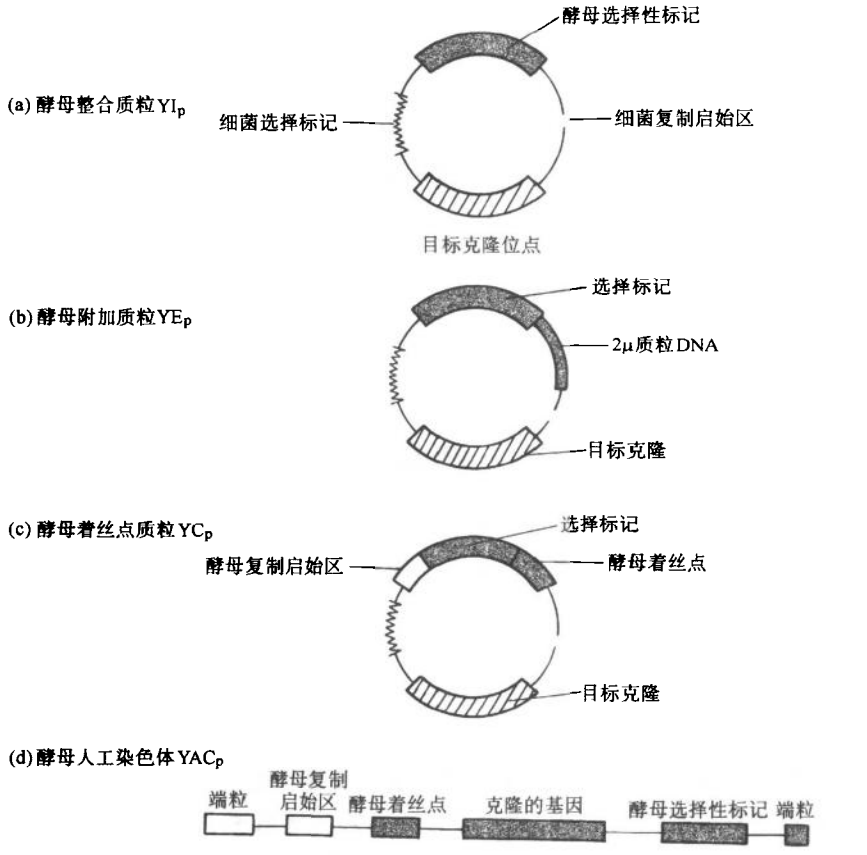


图 9－27 用于酵母的 4 种质粒载体

以被导入大肠杆菌中，因为其上具有能在细菌中复制的起始位点和选择性标记。

能够自主复制的质粒在细胞间传递时，往往会出现子细胞中没有质粒的现象。这是因为质粒在分配过程中是随机的，它依赖于质粒及细胞壁形成的相对位置。然而，如果质粒上包含有一个酵母着丝粒（图 9－27c），那么该质粒将会像染色体一样在细胞分裂时被纺锤体分向两个子细胞中。附加的着丝粒是构成酵母人工染色体的第一步，更进一步是制造一个除包含有一段着丝粒序列以外，在其末端还连有酵母端粒序列的线形结构（图 9－27d）。除此之外，再在该结构中加上一个酵母自主复制序列（autonomous replication sequense，ARS），就构成了一个酵母人工染色体，YAC 在有丝分裂和减数分裂中的行为与酵母染色体是一致的。例如，当一个带有 trp^-ura^+ 的 YAC 和一个带有 trp^+ura^- 的 YAC 的两个单倍体细胞结合在一起形成二倍体后，通过减数分裂形成的四分

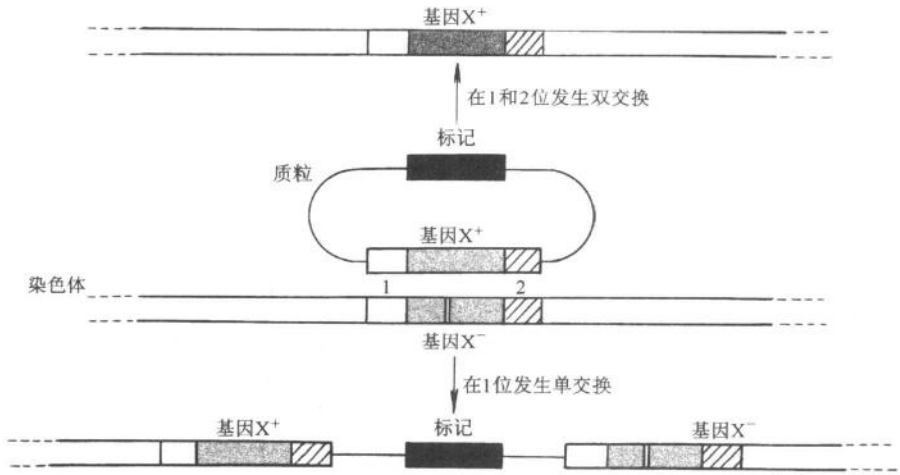


图 9－28　用带有 α^+ 的质粒转化为 α^- 缺陷型酵母受体菌的两种途径

体将如期望的那样出现分离，即两个子囊孢子为 trp^-、ura^+，另两个为 trp^+、ura^-。

二、酵母载体的应用

我们已经知道如何将一个突变基因引入受体细胞并取代其相应的野生型基因。酵母的整合载体（YIP）提供了一个质粒进入真核细胞的模式。一个带有突变基因和与它相邻的序列的整合质粒为同源重组提供了条件，完整的质粒可以插入到酵母染色体的适当位点上（图 9－29）。因此在一条染色体上就同时有一个基因的两个拷贝（一个突变型，一个野生型）。重组体通过染色体环出进行同源配对和交换，可将质粒切掉，这时质粒就带有野生型基因，而突变基因则留在染色体上。

（一）用于对丢失质粒的选择

在某些情况下，我们能够对已经消失的质粒序列进行选择。例如，质粒的选择标记是 lys^+，受体酵母是 lys^-（图 9－29），那么质粒转化的结果是产生带有 $lys2^+$ 和 $lys2^-$ 的菌株。当存在 α－氨基己二酸时，仅有 $lys2^-$ 菌株生长，如果该菌株失去了插入序列，则可以通过在含有 α－氨基己二酸的平板上进行选择。其中大约有一半的菌株保留着原来的野生型基因（x^+），其他则保留了源于质粒的突变型等位基因（X^-）。同时，突变基因 x^- 的表型可用适当的平板直接对其进行选择。

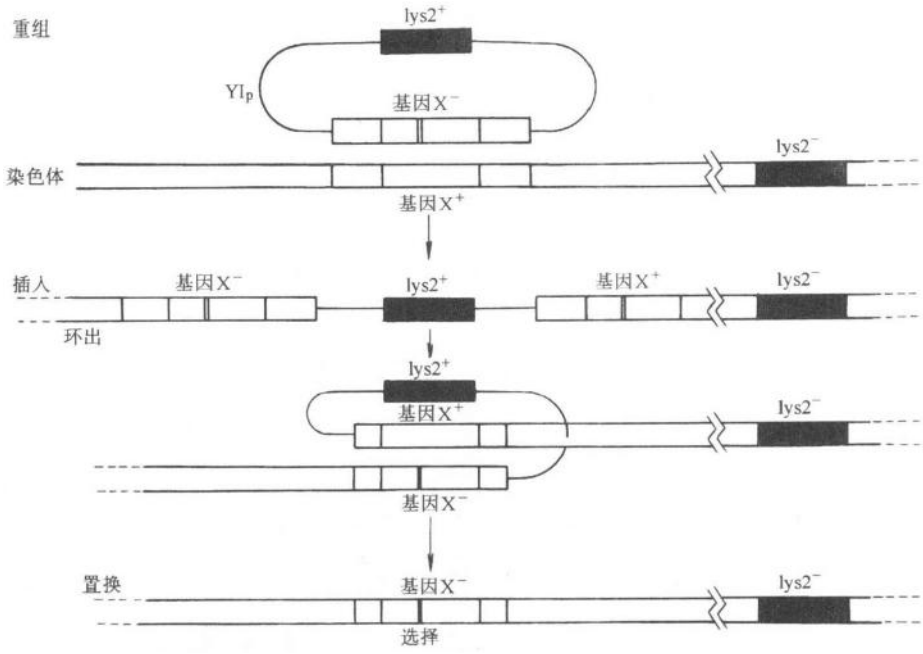


图 9－29　用一个突变基因（X^-）取代一个正常基因（x^+）的两步法

$lys2^+$ 为用于选择的标记基因

前述 Yip 质粒包含两个酵母成分，一个为需要研究的基因 X，另一个为选择标记，两者均能在染色体的特异位点发生同源重组。通过限制酶切割质粒中待研究的基因（X^+），就可以更有效地将靶基因（X^+）插入到染色体的特殊位点上。因为游离末端具有倾向重组的特性，它引导质粒插入到染色体的特定位点上（图 9－30），而两个选择标记基因之间则不会发生重组。

（二）用于基因破坏

有时研究者希望特异性地使靶基因失活，常采用的方法之一是将一具有选择功能的基因插入质粒中待研究基因的野生型等位基因的中部，然后再将这种结构的线状形式特异性地插入到野生型座位上，造成因选择标记基因引起的插入失活，从而确定基因的功能。假定待研究的基因为 $lys2^+$（图 9－31），我们可以将另一基因如 $his3^+$ 插入到质粒的 $lys2^+$ 基因中。然后再以这种结构的线状形式转化基因型为 $lys2^+his3^-$ 的酵母菌，选择 $his3^+$ 转化体。转化体中有一部分为 $lys2^-$，它是由质粒中被破坏的 lys2 等位基因取代了基因组中 $lys2^+$ 所形成的。

（三）研究基因表达的调控序列

着丝粒质粒可用于研究某个基因上游的调控序列（图 9－32）。将基因的

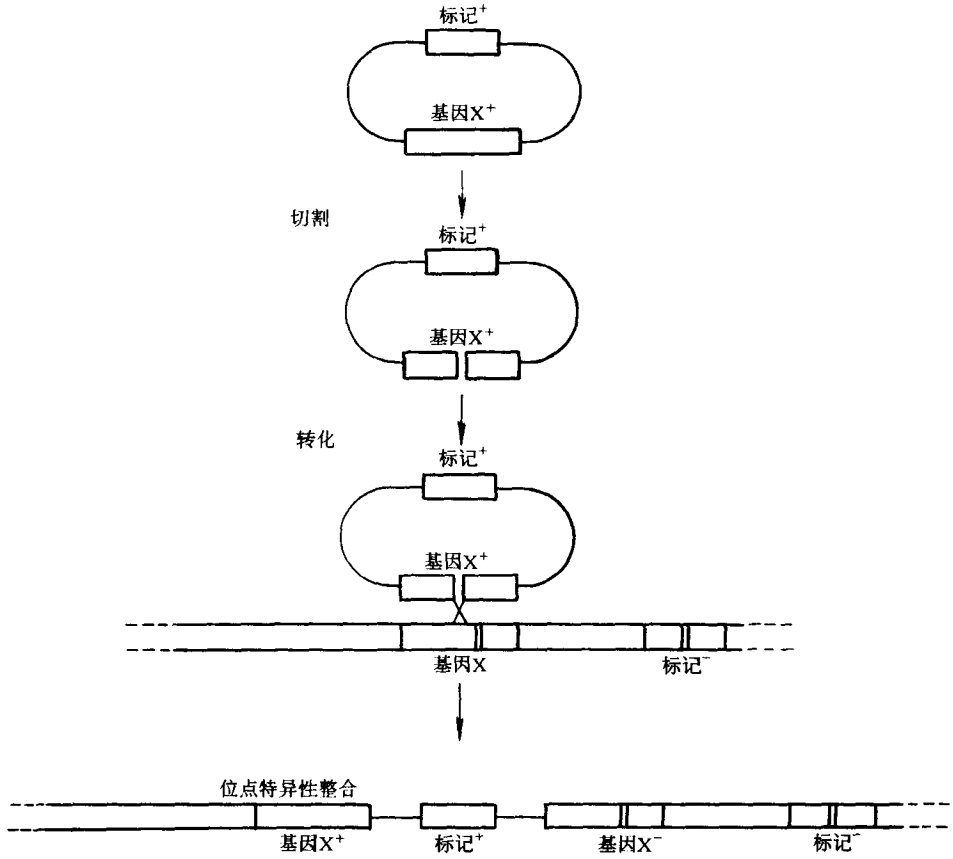


图 9－30　酵母整合质粒通常含一个以上与酵母染色体具同源性的染色体区域
在其适当位置（x^+基因）切割后，可增加其整合的位点特异性

编码区及其上游序列拼接在一起并将其连接到质粒上，这种重组结构可以通过酵母标记如 URA3$^+$加以选择。用特异外切核酸酶在不同程度上从一个方向切割 DNA，诱导上游调控序列产生一系列缺失，随后将其重新连接起来，由此确定哪一个缺失对基因的正常功能没有影响。对某一功能的分析是通过将质粒转入带有缺失突变基因的受体细胞，观察受体细胞中基因功能的恢复情况，以确定对基因的正常功能及表达起调节作用的区域。

在基因表达的调控机制研究中，常用报告基因代替研究的基因。例如在对基因 X 的调控机制进行研究时，要将基因 X 的上游调控区域连接到报告基因上游。一般来说，报告基因具有较易检测的表型，因此基因 X 的正常的调控信号可以通过报告基因来检测。在酵母中运用较多的是细菌 lacZ 基因，它编码 β－半乳糖苷酶。这种酶可以分解乳糖，也能分解乳糖类似物 X－gal。如果

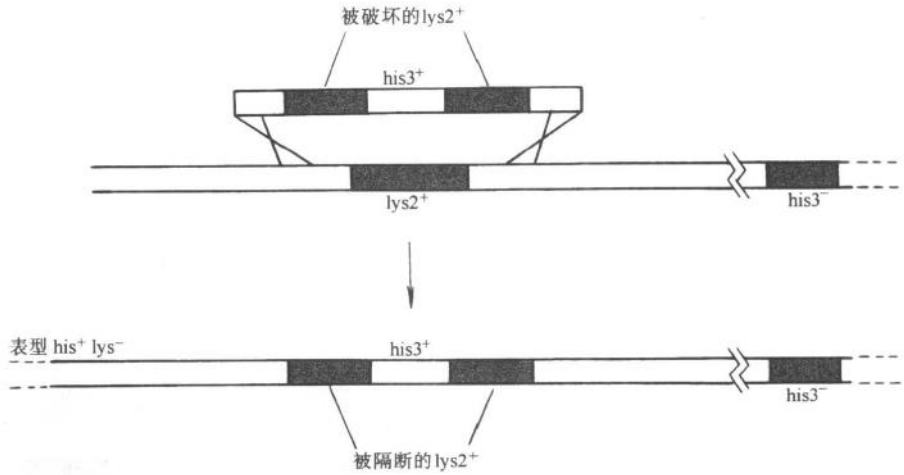


图 9－31　用插入失活的位于线性 DNA 片段上的基因取代野生型基因的一步法

本例中以 his^+ 作为选择转化体的标记基因

培养基中存在 X－gal，酵母细胞在表型上呈兰色克隆。一般而言，重组分子的构建是将基因 X 的调节区及其少数几个密码子按准确的可读框方式连接到编码 β－半乳糖苷酶活性部分的编码区上。

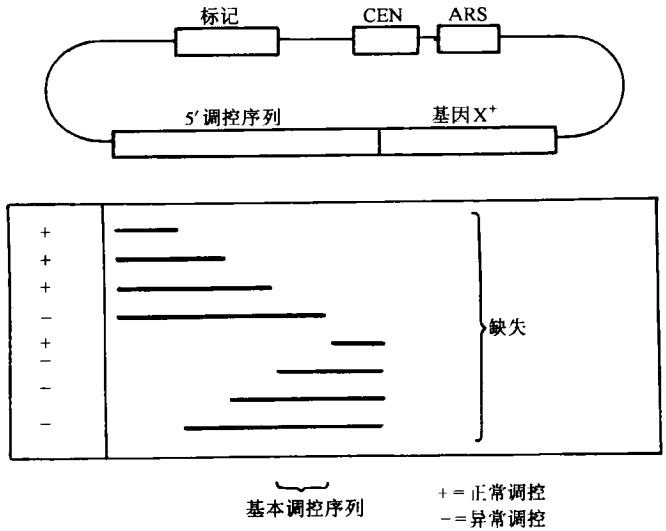


图 9－32　研究酵母基因调控元件功能的策略

首先在离体条件下对基因 X^+ 的调控元件进行缺失分析，然后再将载有不同调控元件的转化结构导入含有缺陷型等位基因 X^- 的酵母菌细胞中，CEN 为着丝粒序列，ARS 为自动复制序列

（四）其他应用

自主复制载体也可用作一种挽回载体（retrieval vector）。如果某个等位基因发生了突变，而且由其产生的表型具有较高利用价值，我们就有必要提取这一发生了突变的基因，研究其结构和功能。在具体操作时，先将带有该基因的野生型等位基因（X^+）的着丝粒载体用适当的限制酶切割，产生一有缺口的载体，然后再用这一有缺口的着丝粒质粒载体转化突变体细胞，质粒就可利用染色体上突变座位的信息来修复缺口，从而将目标座位上的突变序列引入质粒。最后再从酵母细胞中分离已得到修复的质粒，并对其进行分子水平的分析（图 9－33）。

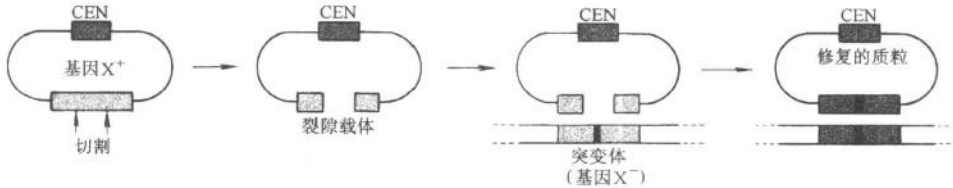


图 9－33　用有缺口的酵母着丝粒质粒提取突变基因的片段

如前所述，酵母的各类人工染色体（YAC）也广泛用于克隆各种真核生物的大片段 DNA。例如编码人凝血因子Ⅷ的基因约有 190kb，而造成假性肥大型肌营养不良（Duchenne muscular dystrophy）的基因则超过 1000kb。如果利用细菌载体克隆人体基因，则需建立一个极其庞大的基因文库。由于 YAC 可以插入很大的片段，则基因文库就会相对小得多。

第四节　植物的遗传转化

通过遗传转化，把控制特定性状的基因转入植物基因组，可以打破原有的种属间生殖障碍，大大缩短育种周期，创造出用常规育种方法很难得到的新型优良品种。遗传转化技术还可以被用来研究基因的结构、功能及其在植物细胞中的表达，它为解决植物生理学、生物化学、分子生物学和遗传学中的许多基本问题提供了一种必要的手段。

一、转基因植物

（一）概述

转化的主要目的是让外源 DNA 稳定地插入植物细胞的染色体组中，并再生出完整的植株。所以，作为转化受体应符合两个条件：（1）能够接受外源基

因，并通过基因重组或其他途径使外源基因稳定地插入植物染色体组中；(2)必须具有脱分化和再生能力，能够形成新的植物体。在许多植物中，同时符合这两个条件的受体细胞是很难找到的。植物遗传转化技术的发展历程正是人们通过技术改进和创新，不断扩大受体细胞的类别，提高转化细胞或组织的再生能力，从而增加适于遗传操作的植物种类，简化转化程序，提高转化效率的过程。目前人们经常所用的几种植物转化组织或受体细胞见表9－6。

表9－6 几种植物转化受体组织或细胞

细胞类型	转化方法
叶片和茎段等（包括子叶和胚轴）	农杆菌介导，基因枪等
原生质体	电击法、PEG、脂质体、微注射等
愈伤组织和悬浮培养细胞	基因枪、超声波等
未成熟胚和分生组织	农杆菌介导、基因枪、超声波等
花粉	基因枪、浸泡法等
子房或胚珠	花粉管通道、子房注射等

用于真核细胞转化的载体具有一些共同特性。如它们都有细菌的复制起始位点（如ColE1），在原核生物中表达的选择标记基因（如氨苄青霉素、卡那霉素等抗性基因）和多克隆位点（MCS）等，这些结构使质粒能够在大肠杆菌中保留并大量复制，同时也有利于各种DNA重组操作和插入不同的目的基因。植物转化载体的另外一个重要组成是能够在植物细胞中表达的标记基因，它对于转化体的识别和选择是必不可少的。标记基因又分为两类，选择标记基因和报告基因。目前在植物遗传转化研究中常用的选择标记基因见表9－7。报告基因大多编码一些易于检测的酶，如氯霉素乙酰转移酶、葡糖苷酸酶、荧光素酶、胭脂碱和章鱼碱合成酶等。迄今为止，卡那霉素是使用最广泛的选择剂。除草剂的应用也越来越受到人们的重视。在几种报告基因中，GUS基因虽使用较晚，但由于其灵敏度高，易于检测等特点，已成为目前植物遗传转化研究中应用最广泛的报告基因。

表9－7 常用选择标记基因

基因	编码产物	选择剂	来源
抗生素			
NPT	新霉素磷酸转移酶Ⅱ	卡那霉素，G418	Bevan 等（1983）
HPT	潮霉素磷酸转移酶	潮霉素	Van den Elzen（1985）
DHFR	二氢叶酸还原酶	氨甲蝶呤	Herrera－Estrella 等（1983）

（续）

基　因	编码产物	选择剂	来　源
Ble	不详	博莱霉素	Hille 等（1986）
除草剂			
als	乙酰乳酸合成酶的变异型	Chlorsulfuron	Haughn 等（1986）
aro A	5－烯醇丙酮基莽草酸 3－磷酸转移酶	草甘膦	Comai 等（1985）
bar	膦丝菌素转乙酰基酶	膦丝菌素	De Block 等（1987）

标记基因大多是从微生物中分离出来的，只有与在植物细胞中能够表达的启动子连接才能起作用。目前使用最多的启动子是从花椰菜花叶病毒（CaMV）中分离出来的35S启动子。其次是从农杆菌（*Agrobacterium tumefaciens*）Ti质粒中分离出的胭脂碱和章鱼碱合成酶的启动子（NOS和OCS）。它们具有真核生物启动子的一些特点，能在植物细胞中实现基因表达，且都是组成型表达，没有组织和发育特异性。近两年来，人们已经从植物染色体组中分离出一些具有组织或发育特异性的启动子，如在花药组织中特异表达的启动子，受伤后表达的启动子和在单子叶植物水稻中分离的肌动蛋白启动子等。

（二）植物转基因方法

1. 农杆菌介导的基因转移　根癌农杆菌（*A. tumefaciens*）是一种土壤细菌，能够感染许多双子叶植物的受伤组织，引起冠瘿瘤（crown gall）。在根癌农杆菌中存在着一种致瘤质粒（简称Ti质粒），该质粒上有一段DNA序列（T－DNA）可以转移到植物基因组中。人们正是基于这一原理对Ti质粒进行改造，将目的基因转入质粒的T－DNA区，构建农杆菌的转化载体。常用的农杆菌转化载体有共整合载体（cointegrate vector）和二元载体（binary vector）。

（1）共整合载体　所谓共整合载体即将外源基因插入到一个中间载体上，然后中间载体通过与除掉了致瘤基因的Ti质粒结合形成共整合体（图9－34），最后用于转化植物的载体系统。这种载体系统首先由斯切尔等（Schell等，1983、1984）设计构建，由两个质粒组成。一个为非致瘤性Ti质粒，这种质粒除掉了致瘤基因，保留有边界序列和Vir区域并且还在两个边界序列之间接有一个拷贝的pBR322，这就是所谓的接收载体（receptive vector）如pGV2260。第二个质粒为装有植物转化体选择标记基因如nptⅡ基因的pBR322质粒，异即中间载体（intermediate vector）。由于这两种质粒都含有pBR322序列，所以在农杆菌细胞内就可以通过同源重组形成共整合体。

当目的基因装到中间载体上以后就可以转化大肠杆菌，然后再从大肠杆菌

转移到农杆菌内形成共整合体。由于 pBR322 无自动转移功能，必须有第二种质粒的帮助，中间载体才能从大肠杆菌转移到农杆菌中。如质粒 pR64drdⅡ能以反式互补的方式提供转移功能。其次细菌结合需要质粒移动识别基因即 mob 基因，mob 基因的编码产物能识别一种特殊的序列 bom 位点。pBR322 具有 bom 位点但不含有 mob 基因，因此还需要第三种质粒如 pGJ28 以反式互补的方式提供 mob 功能。这三种质粒相互亲和，能共处于同一大肠杆菌细胞内，因此在供体大肠杆菌细胞内存在三种质粒。这三种质粒转移到农杆菌细胞中以后都不能复制，中间载体只有与接受载体 pGV2260 结合以后才能保存下来。

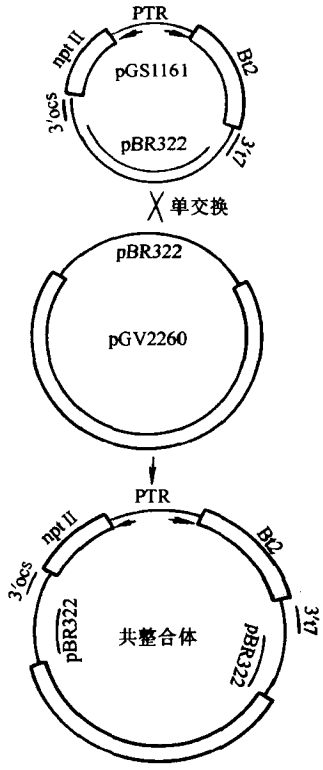


图 9－34　共整合载体系统
PTR 为 TR－DNA 的启动子（482bp，箭头示），3′t_7，为 TL－DNA 基因 7 的多聚 A 信号序列（211bp），3′ocs 为章鱼碱合成酶基因的多聚 A 信号序列（706bp）Bt2 为苏云金杆菌的毒素蛋白基因（根据 Vaeck 等（1987）重绘）

（2）二元载体　当 Ti 质粒的 T－DNA 和 Vir 区域分别位于两个独立的复制子上时，Vir 基因能够以反式互补的方式帮助 T－DNA 整合到植物染色体上，这就是所谓的二元载体系统。二元载体的主要优点是操作简便，不需要构建质粒共整合体，即无需通过相互整合的方式将基因转移载体结合到 Ti 质粒上。

霍克曼等（Hoekeman 等，1983）首先报道二元载体系统。例如农杆菌 LBA4404 含有质粒 pLBA4404，这种质粒是章鱼碱 Ti 质粒 pTiAch5 的衍生质粒，具有 Ti 质粒的 Vir 区域但缺失了 T－DNA 部分。利用 pLBA4404 作为帮助载体（helper vector）使另一载体上 T－DNA 的两个边界序列之间的所有序列整合到寄主植物的染色体上。

艾当等（Adang 等，1987）利用二元载体系统也成功地将苏云金杆菌中的杀虫晶体蛋白质基因（ICP）转移到烟草中，也培育出烟草杀虫植株。所使用的帮助载体也是 pLBA4404，但是基因转移载体为 pH400 如图 9－35。质粒 pH400 具有章鱼碱合成酶基因（ocs）、25bp 的边界序列（A 和 B）和一个转化体选择标记基因。转化体选择标记基因由 CaMV 病毒的 19S 基因的启动子、nptⅡ基因的编码序列和加尾信号序列三部分组成。ICP 基因也为融合基因，

即 ICP 编码序列的 5′端接有 T－DNA 启动子 24，3′端接有多聚 A 加尾信号。目的基因装在载体 pH400 上组成质粒 pH450 以后，再转移到农杆菌 LBA4404（pLBA4404）中，最后用于转化烟草叶组织。由于质粒 pH450 中所有需要转移的基因都位于两个边界序列 A 和 B 之间，因此 Vir 基因就可通过反式互补，使其转移到植物染色体上。

农杆菌是通过伤口侵入的。所以，所有接种方法必须事先在受体细胞或组织上造成伤口。目前常用的农杆菌转化方法有叶圆片法、共培养法和直接接种法等。

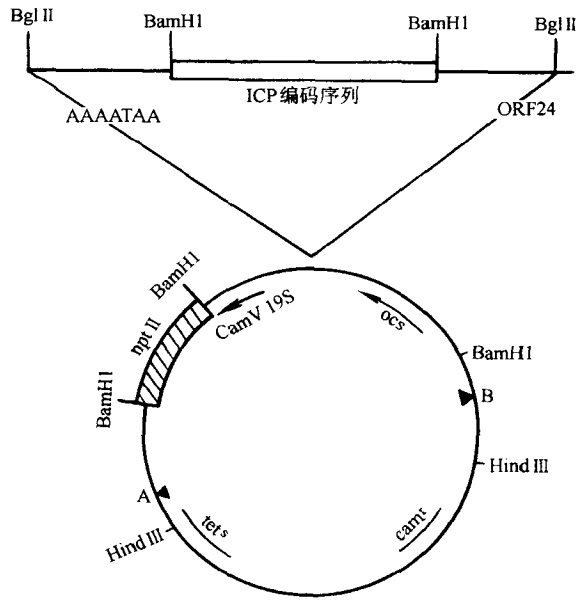


图 9－35　二元载体系统中基因转移载体及重组杀虫蛋白质基因

CaMV19S 为 CaMV19S 基因的启动子，ocs 为章鱼碱合成酶基因，A 和 B 为 T－DNA 的边界序列，tet^s 和 cam^r 分别为四环素敏感和氯霉素抗性标记基因，ORT_{24}为 T－DNA 的启动子 24（根据 Adong 等，1987）

2．基因枪转化系统

基因枪最初是由 Sanford 等人（1987）设计的，它利用火药爆炸或其他动力加速包裹了 DNA 溶液的金属粒子（如钨、金等），在金属粒子射入植物细胞时也把 DNA（RNA 或其他生物大分子物质）带入植物细胞。与其他转化方法相比，基因枪法具有如下一些优点：1）克服了农杆菌转化的寄主限制。在基因枪转化过程中 DNA 是机械导入的，从理论上讲不存在任何寄主限制。特别是农杆菌不能侵染的重要禾谷类粮食作物玉米、小麦、水稻和燕麦等，用微弹轰击法都先后获得了转化的再生植株。2）简化了质粒构建。目的基因在大肠杆菌中直接克隆、扩增，从而可以去掉在农杆菌中复制和转移所必需的 DNA 序列。基因枪轰击可以用两个或两个以上质粒进行共转化，而不需要把不同的目的基因克隆到一个大质粒中。3）简化了转化方法。利用原生质体建立的转化系统，如电击法、PEG 诱导法和脂质体转化等都需要建立良好的原生质体培养再生系统，这一工作繁琐、费时，而且在很大程度上受基因型的影响。基因枪轰击方法简便，易于建立适合于所有品种的转化系统。4）可避免农杆菌介导的转化程序中因农杆菌污染造成的假阳性。

3. 其他转化方法　去壁后的裸细胞，也就是原生质体可以通过细胞吞噬作用 吸收外源大分子物质，从而有可能成为基因转移的良好受体材料。转化原生质体一般利用大肠杆菌的小质粒作为载体。从理论上讲，原核生物载体只需经过少量改造，也就是加入能在植物细胞中表达的标记基因就可以在 DNA 直接转移中加以利用。许多在农杆菌质粒改造中使用的中间载体和二元载体都已被成功地用于 DNA 直接转移法。

目前，通过植物原生质体系统进行转化有许多不同的方法，如电击法、聚乙二醇（PEG）介导法和质脂体介导法等。

二、植物基因操作的研究方向与重要基因的鉴定

利用基因工程技术可以把控制特定性状的基因转入现有植物品种，大大缩短育种周期，创造出用常规育种方法很难得到的新型优良品种，因此在农作物品种改良中展现了广阔的应用前景。改良农作物生产的最直接的办法，可能就是应用基因工程技术改变作物的某些遗传特性，使它们获得新的丰产优质的遗传性状。据不完全统计，目前已获得转基因植株的物种已达 70 多种，其中包括重要的粮食作物如水稻、小麦、玉米、大豆和马铃薯等。基因工程学家们已经获得一些抗除草剂的转基因植物，包括棉花、大豆、烟草和西红柿。当然如果这类植物得到大面积的种植，无疑会得到经济上的好处。但这是一项十分复杂的工作，因为新的除草剂还会不断涌现，从而使已有抗性的植物失去作用。而且对我们国家来说，培育抗除草剂的转基因植物并不一定是解决农业生产的迫切问题。

（一）光合作用效率的遗传改良的可能性

绿色植物光合作用的产物，约占植物干重的 95% 以上，它是地球上一切动物，也包括大多数微生物的生命源泉，同时也是人类社会的主要生活物质和能量的来源。然而在目前，地球上的植物利用太阳能的效率相当低，因此如何有效地提高光合作用的效率已成为研究植物基因工程活跃的领域。

在光合作用的 CO_2 固定过程中起关键作用的酶是二磷酸核酮糖羧化酶（RuBisco）。现在许多科学家将 RuBisco 作为基因工程的目标之一，以期通过提高 RuBisco 同 CO_2 的亲和力来改善光合作用的效率。RuBisco 是由 16 个亚基组成的复合体，包括 8 个大亚基（rbc－L）和 8 个小亚基（rbc－S），前者是由叶绿体基因组编码的，后者是由核基因编码的。通过将不同来源的 RuBisco 的基因导入同一植物，形成具有异源亚基的 RuBisco 基因；或采用定点突变技术，改变 RuBisco 的活性，增加其同 CO_2 的亲和力；甚至用更为有效的突变基

因取代正常的 RuBisco 基因等办法，将有可能提高植物对 CO_2 的固定效率。

叶绿体的基因操作目前也引起了普遍关注，人们设想通过改变光合系统 I 和光合系统 II 反应的中心蛋白质编码基因 psaA 和 psbA 来改变植物的转能效率，使其达到最佳水平，从而提高光合作用的效率。

最近几年，在叶绿体的基因操作方面已经取得了多方面的进展。已完成了对烟草和水稻等植物的叶绿体基因组 DNA 的全序列分析；成功地从小麦、水稻、玉米等多种植物中克隆到了若干个光合系统的基因。几种光诱导的基因的表达调节机制也正在分析之中，等等。

（二）固氮的遗传改良

在过去的 20 多年中，关于固氮根瘤菌同豆科植物之间的遗传共生问题的研究已取得了很大的进展。根瘤菌属（*Rhizobium*）的细菌侵染豆科植物的根部，与豆科植物建立共生关系，在根部形成根瘤，才能够进行正常的固氮作用。固氮作用是在根瘤中的固氮复合酶的作用下完成的。固氮复合酶可以将 N_2 还原为可被利用的 NH_4^+。根瘤的发生是由细菌的根瘤基因（nod gene）控制的。在根瘤内部，根瘤细菌发育成为能够固氮的多型“类菌体”。固氮过程既需要细菌蛋白，也需要植物蛋白。细菌蛋白由 nif 基因编码，构成了氧敏感的固氮复合酶和电子传递链的组分。

把非豆科植物，特别是禾谷类农作物转变为固氮植物，这是植物固氮基因工程的一个重要目标。一种可能的途径是用带有 nif 基因的质粒转化植物细胞的叶绿体，从而可能使用正常的原核信号进行表达。但这种方法的困难较多。另一种看来是最有效的方法是，把豆科植物的根瘤基因导入其他植物细胞，使之能够对固氮的感染作出恰当的反应。这项工作必须对影响固氮菌与植物根部相互作用的机制有深刻的了解。到目前为止，许多植物的根瘤蛋白基因（nodulin gene）都已经被克隆出来了。

（三）植物种子贮藏蛋白的改良

禾谷类植物和豆科植物的种子含有丰富的贮存蛋白，是人类重要的蛋白质来源。世界上大约有 70% 以上的蛋白质来源于禾谷类，20% 来源于豆类。已知人类有 10 种必需氨基酸是人类自己无法合成的，得从外界供应的食物中获得。而禾谷类和豆类的种子贮存蛋白又恰恰缺少某些类型的必需氨基酸。科学家设想的克服种子营养缺陷的途径是：应用重组 DNA 技术改变种子贮存蛋白基因的特定碱基序列结构，使之加上缺少的必需氨基酸的密码子。现在有关的一些种子的贮存蛋白的基因包括菜豆蛋白基因、豆球蛋白基因和玉米醇溶蛋白基因都已相继得到了克隆。因此对植物种子贮存蛋白的改良将会得到实现。

（四）培育抗病毒的农作物

病毒感染使许多农作物的产量显著下降，给农业生产造成巨大的损失。据统计在美国西部，每年因西红柿花叶病毒（TMV）感染而造成的直接经济损失达5000万美元。防止TMV病毒感染的一种有效的办法是，给植株注射TMV弱病毒株，以防烈性TMV毒株的严重感染，为植物提供交叉保护。这就是所谓的诱导抗性。许多植物具有广谱抗性，但它必须在受到病原菌的感染之后，才能诱发产生。植物的这种“诱导抗性”现象与哺乳动物接种疫苗之后所呈现的免疫反应十分类似。给植物接种上弱毒的毒株，或解体的病原菌的各种组分，就能诱发产生出广谱抗性。在细菌、真菌及类病毒的感染中，也都观察到了这种现象。此种通过人工注射法的植保手段，十分耗费资金和劳力。理想的办法是应用基因工程技术，从自然界中鉴定出抗病毒的基因，然后通过重组转化等手段，培育出携带并能遗传抗病基因的转基因植物。

要培育抗病毒的转基因植物，首先得加强有关病毒感染的生化、分子遗传学等方面的基础研究。目前我们对于植物抗病的分子机理以及交叉保护的生化基础还很不了解。但有一种模式认为，预感染病毒的外壳蛋白质是决定交叉保护反应的重要因素。这种思想构成了当前设计抗病毒转基因植物的理论依据。有人已应用Ti质粒载体系统和叶盘转化技术，成功地将西红柿花叶病毒外壳蛋白质（CP）的cDNA插入到西红柿植株上，其中CP－cDNA是使用CaMV－35S启动子进行表达的。根据卡那霉素抗性选择出来转化细胞，再由此生成成熟的转基因植株。这些植株含有1～5个拷贝的CP－cDNA，而且从叶片提取物中测得的TMV－CP蛋白质占叶片总蛋白的0.1%。用TMV感染此种转基因西红柿的籽苗，试验组植株疾病症状的表现明显晚于对照组的植株。同时还有一定比例的试验植株（依不同的转基因品系而定，范围约4%～47%）并不出现任何感染的症状。

N.E.Tumer等人（1987）把苜蓿花叶病毒外壳蛋白质基因导入烟草和西红柿，也获得了与此同样的结果。培育抗病毒转基因植物的下一步目标可能是小麦、玉米和大麦等重要粮食作物。因为据有关资料表明，小麦病毒、玉米花叶病毒和大麦花叶病毒每年造成作物减产，经济损失也是相当严重的。国内近几年来先后克隆了烟草花叶病毒（TMV）、黄瓜花叶病毒（CMV）、玉米矮花叶病毒（MDMV）、马铃薯X病毒（PVX）和马铃薯Y病毒（PVY）等的外壳蛋白基因，以及烟草花叶病毒的卫星RNA，黄瓜花叶病毒和马铃薯Y病毒的复制酶基因等，获得了抗TMV和CMV的转基因烟草，并已进入大田试验。抗CMV的番茄也已在温室中试种。特别是最近我国自主克隆了小麦的重要病害－大麦黄矮病毒（BYDV）的外壳蛋白基因，并转入小麦，转基因植株对大麦黄矮病毒的GPV株系具有高度抗性。

（五）培育抗虫害的农作物

植物的虫害是造成农作物减产的最严重的自然灾害之一。各国政府及有关的农业科研机构历来都十分重视开展以此为目标的植物保护研究工作。据估计，世界每年因有害生物危害而损失粮食总产量的35%以上。在我国，水稻因虫害减产10%以上，小麦减产近20%，棉花产量损失达20%～30%。粮食、棉花和水果等农产品在储藏过程中还要受到许多病虫的危害，也造成相当大的损失。多年来，病虫害的防治主要靠化学药剂，在生产上收到积极的效果。随着时间的推移，特别是化学农药的不合理使用，引起一系列副作用，如诱导害虫产生抗药性，天敌被杀灭，自然生态环境被破坏，引起次要害虫的大发生及环境污染，残毒增加等。这些都成为农业生产上迫切需要解决的问题之一。实践证明，培育抗病虫害的抗性农作物新品种是一种既经济实惠，又无不良副作用的有效途径。因此，应用基因操作技术，创造出具有抗病虫害的农作物新品种，具有明显的经济效益。

在抗虫作物育种方面一条可能的途径是直接将细菌毒素基因导入敏感植物，使之获得抗害虫的遗传特性。例如苏云金芽孢杆菌（*Bacillus thuringiensis*）所产生的δ－内毒素蛋白质（Bt毒素），对许多种农业害虫的幼虫都具有高度特异的致死效应。Bt毒素已经被纯化出来，用它喷洒在植物的叶子上，也可以达到杀灭害虫的效果。但这种方法费用过于昂贵，无法在农业生产上大规模推广使用。通过基因重组和转基因技术，可以把Bt毒素基因转移到敏感的作物中去，使之对农业虫害具有内在抗性。据报道，应用这种技术已经培育出了一批能够表达Bt毒素的烟草、马铃薯、西红柿、棉花等植物。它们对玉米穗蛾、棉铃虫、甜菜夜蛾以及烟草夜蛾等害虫都具有杀虫的效力。目前的研究方向之一是继续寻找更好的毒素基因。

1996年，转Bt基因的棉花、玉米和马铃薯已在美国批准进行商品化生产。国内虽早已克隆了Bt毒蛋白基因，并已转入烟草、玉米、杨树等植物中，但仍需要对Bt基因的密码子进行改造，使之具有杀多种害虫的功能。目前，转基因抗虫棉已进入大田示范种植，预计在21世纪初抗虫的转基因棉花、玉米、大豆、水稻、白菜等作物将在我国大面积推广。除了Bt毒素基因外，还有一些抗虫基因在农业上也得到应用。昆虫体内存在许多水解蛋白酶，如胰蛋白酶、胰凝乳蛋白酶等丝氨酸类蛋白酶，它们是昆虫生理代谢过程中所必需的酶，负责裂解和消化其食物中的蛋白质。当昆虫体内这些酶受到抑制时，它们就不能消化食物中的蛋白质，饥饿而死。蛋白酶抑制剂是自然界最为丰富的蛋白质之一，存在于绝大多数生物体尤其是许多植物储藏器官如种子和块茎中，各种蛋白酶抑制剂占总蛋白质的1%～10%，有些果实中丝氨酸蛋白酶抑制剂

占总蛋白量的50%。与前述的苏云金杆菌毒素相比，蛋白酶抑制剂具有抗虫谱广，昆虫不易产生耐受性等优点。到目前为止，已从大豆、番茄、马铃薯、大麦等植物中分离克隆出多个丝氨酸蛋白酶抑制剂基因，发现豇豆胰蛋白酶抑制剂（cowpea trypsin inhibitor，CpTI）抗虫效果最为理想。CpTI是由一个小的多基因家族所编码的有80个氨基酸的小分子多肽，它在豇豆种子成熟时大量积累。由于CpTI的抑制作用的位点是酶活中心，除非其消化系统主要生理生化过程都发生变化，否则害虫几乎不可能通过突变的方式诱发抗性。所以，CpTI所表现的抗虫性是比较稳定的。

杆状病毒能引起昆虫的流行病，是效果很好的微生物杀虫剂，但由于其寄主范围窄，致死速度慢而在实际应用上受到很大的限制。有科学家将氯霉素乙酰转移酶基因插入苜蓿夜蛾核多角体病毒（AcNPV）基因组中，在广泛启动子RSV－LTR的控制下，AcNPV所携带的目的基因可以在除鳞翅目外的其他昆虫如蚊、蝇中表达，说明重组病毒的有效寄主范围扩大了。如果同时用重组DNA技术将病毒的早期启动子与对昆虫有毒杀致死作用的基因（如Bt内毒素基因、CpTI基因及神经肽基因等）相连，然后插入病毒基因组或置换基因组中的非必要部分，就能产生迅速杀死昆虫的重组病毒。

（六）培育抗除草剂的农作物

抗除草剂基因工程植物在国外很受重视，而我国在这方面的研究相对较为薄弱。随着经济的不断发展和人工费用的提高，今后除草剂的使用将越来越普遍。

除草剂是一类用于清除大田杂草的化学药剂，但它往往也会抑制作物的生长。因此，所施用的除草剂对杂草必须具有选择致死效应，而且要严格地控制喷洒的地域，这样才能真正取代机械操作法，达到除草保苗的目的。但如果作物本身具备了抗除草剂的特性，除草剂的使用自然就会更加方便有效。如何创造出抗除草剂的作物新品种，已成为当前植物基因工程的主要目标之一。

自从20世纪50年代发现了脲素型的除草剂（urea type herbieide）以来，关于除草剂的研究发展十分迅速。今天，除草剂的应用已成为现代化农业生产中不可缺少的重要环节。化学除草剂不仅可以用于农业生产，而且也可以用于军事目的，使敌方的农作物致死。除草剂不仅有实用价值，而且在基础理论研究方面也有重要的意义。

就目前所知，现在使用的除草剂主要分两大类：一类通过破坏氨基酸合成途径杀死杂草，如草甘膦。草甘膦是一种高效广谱的灭生性除草剂，它具有易降解，不污染环境等突出优点，其缺点是选择性差，常常对许多农作物造成伤害。它是通过抑制芳香族氨基酸生物合成中起关键作用的EPSP合成酶而破坏

植物的正常生长。1987 年，美国科学家成功地从矮牵牛中克隆出 EPSP 合成酶基因，接上 CaMV－35S 启动子以后，转入油菜细胞的叶绿体中，转基因油菜叶绿体中 EPSP 合成酶的活性大大提高，并且能够有效地抵抗草甘膦的毒杀作用。另外有人把降解除草剂的蛋白质编码基因导入宿主植物，从而保证宿主植物免受其害。该方法已被成功地用于选育抗磷酸麦黄酮的工程植物。磷酸麦黄酮是谷氨酸类似物，它作为除草剂的一种有效成分破坏了谷氨酰胺合成酶的正常作用，从而阻断氨基酸的生物合成。比利时科学家将乙酰转移酶基因导入烟草、番茄和马铃薯中发现，转基因植物可以在相当高的磷酸麦黄酮溶液中生长，主要是因为该化学物质进入植物细胞中后立即被乙酰化而失去功能。另一类是通过破坏植物光和作用中电子传递蛋白质的功能来杀死杂草。一种广泛使用的三氮杂苯类除草剂莠去净（atrazine）就是由于它能够特异地同光系统 II 反应中心蛋白质 Qβ 结合；从而阻断了高能电子的转移，最终导致植物死亡。莠去净的受体蛋白 Qβ 是由叶绿体 DNA 的 psbA 基因编码的。但有一些杂草（如苋菜）对莠去净的作用具有天然抗性。经过基因克隆和 DNA 序列结构分析发现。抗性型苋菜的 psbA 基因同敏感型的相比，仅仅是在第 264 位的一个氨基酸密码子发生了变化（Ser→Gly）。然而，在高等植物中，这种单一氨基酸的取代反应就会阻止 Qβ 蛋白质同除草剂莠去净的结合，使植物具备抗性功能，却不会影响光合作用的正常进行。此外，在一些原生生物和蓝细菌（cyanobacterum）中，Qβ 蛋白质第 264 位上也存在着单一氨基酸取代反应（Ser－Ala），它们也具有抗除草剂阿特拉津的能力。这些研究都为将抗除草剂基因引入主要的农作物如大豆、玉米、水稻、小麦等奠定了理论基础，创造了有利的条件。

（七）培育抗真菌、细菌病害的作物

我国科学家把抗菌肽基因转入烟草和马铃薯，转基因植株对细菌性病害－青枯病产生抗性，抗性比起始品种提高 1～3 级。将几丁质酶基因转入植物的研究也已起步。总之，抗病基因工程的研究才刚刚开始，在 21 世纪必定会成为植物基因工程研究的一个热点。

（八）增加植物次生代谢产物的产率

植物次生代谢产物为人类社会提供了大量的有用商品，包括各种药物、酶制剂、生化试剂、化装用品以及食品添加剂等。据统计，世界上的药物供应约有 25％左右是直接或间接地来自植物。例如治疗疟疾的特效药奎宁、治疗白血病的长春新碱、避孕药的重要原料薯蓣碱、治高血压的东莨菪碱、麻醉剂吗啡因、可待因以及强心剂地高辛等。至于中草药中使用的植物药材更是难以胜数。最近，又从玉米或一些生长在非洲的植物中分离到甜味剂，著名的如高果

糖玉米甜味剂（high－fructose corn sweetner，HFCS）比通常的食糖要甜几千倍到十几万倍。以上这些产品，大都是从植物中直接纯化而来的。但根据目前的技术发展水平，我们有可能通过基因工程方法提高植物次生产物的合成效率，而且有可能通过基因操作，由植物生产出具有优良特性的全新的化学物质。当然这是一项较为长远的目标。在着手进行这项工作之前，需要进行广泛深入的有关植物分子生物学和生物化学的基础研究。另外，植物基因工程技术在果实延熟保鲜、创造雄性不育、作物品质改良，提高抗逆性（抗寒、抗盐等）、生物反应器生产外源物质等方面的研究也将有一个大的发展。

（九）用植物基因工程的方法获得雄性不育株系

1．*用特异表达细胞毒素获得雄性不育植株* 将花粉或花药特异启动子、细胞毒素基因以及转录终止子3部分组装构建成嵌合基因并转化植物，细胞毒素的特异时空表达能够选择性地破坏与花粉发育相关的某些器官或组织，阻断其发育进程，导致植物雄性不育。有人将烟草花药绒毡层特异表达基因TA29的启动子分别与核糖核酸酶T1和芽孢杆菌RNA酶（Barnase）基因融合，构建表达盒TA29－Rnase T1和TA29－Barnase，并通过农杆菌的双元载体系统转化烟草，发现绝大多数（92%）带有TA29－Barnase基因的植株不能产生正常的种子，但若以这些植株作母本，用其他植株的花粉进行人工授粉，就能产生发育正常的种子。表明这些TA29－Barnase转化株是雄性不育，但雌性可育。含单拷贝的TA29－Barnase雄性不育烟草做母本与正常的烟草杂交的F1代植株，一半为雄性不育，一半为雄性可育株。

2．*用反义RNA技术获得雄性不育植株* 花粉发育是一个极其复杂的过程，许多基因与花粉发育有关，通过反义RNA技术阻断与花粉发育有关基因的表达同样可以获得雄性不育植株。类黄酮是花粉发育过程中的重要物质，而苯基乙烯酮合成酶（CHS）是其生物合成的关键酶。有人将chs基因的反义RNA基因与CaMV－35S启动子及花药特异序列（anther box）串联构建成嵌合基因，导入矮牵牛发现，由于chs反义RNA在花药绒毡层的表达，抑制了Chs的合成，阻碍花粉的正常发育，从而得到了雄性不育的矮牵牛。这些转基因植物花冠颜色变浅，花药变白，花粉粒异常或不萌发形成花粉管。由于植株雄性不育，而雌性正常可育，说明chs反义RNA的表达对雌蕊发育无明显不良影响。

3．*通过提早降解胼胝质壁获得雄性不育植株* 在花粉的发育过程中，小孢子的分离要靠花药绒毡层分泌胼胝质酶降解坚硬的胼胝质。这个发育过程有着严格的时间性，如果提前分泌胼胝质酶，过早降解胼胝质，小孢子发育就会停止，产生畸形花粉粒，导致植株雄性不育。有人将与致病性有关的β－1，3－

葡聚糖酶基因分别与拟南芥菜花粉发育早期启动子 A9 或 A3 融合，转化矮牵牛后获得不同程度的雄性不育植株。不育植株的花药小，绒毡层大，无花粉粒，胼胝质壁在减数分裂的早期即被降解，而小孢子从四分体中释放不久即破裂。

4. 其他策略　有人将发根农杆菌的 rolC 基因与 CaMV－35S 启动子串联成嵌合基因，研究转基因烟草发现，rolC 的系统表达，使得植株在形态上发生了一系列变化，包括株高下降，顶端优势减弱，叶片色素水平下降及雄性不育，还导致花药中 IAA 含量增加，GA 含量下降，估计这类基因有可能通过破坏花粉细胞中的激素平衡引起花粉败育。有些胞质型雄性不育与其线粒体所带基因有关，通过基因工程的方法扰乱线粒体与细胞核之间遗传信息的相互作用，也能导致植株雄性不育。

5. 雄性不育的保持与恢复　用基因工程雄性不育系做母本，用正常植株做父本杂交，其后代往往出现分离现象。所以，如何使雄性不育株的后代完全雄性不育是该技术应用于农业生产的关键。将抗除草剂基因与雄性不育基因串联在一起，构建用于转化植物的表达载体，有可能解决雄性不育系的保持问题。在这样的转基因植株中，不育基因与抗除草剂基因是紧密连锁的，两者同时存在或同时消失，我们可以通过施用除草剂有选择性地杀死雄性可育植株，而将不育株保持下来。

一些雄性不育植物本身编码导致雄性不育的物质，如不育 RNA，不育蛋白或多肽等，可以将这些不育物质的反义 RNA 基因或其蛋白质及多肽的抑制剂基因导入相应的正常植物，以求获得其恢复系，这个方法不仅可以为自然的雄性不育创造恢复系，也可以为基因工程雄性不育植物创造恢复系。

有科学家将芽孢杆菌 RNA 酶的抑制剂 Barstar 基因与烟草花药绒毡层特异启动子 TA29 串联成嵌合基因 TA29－Barstar，导入油菜并得到 TA29－Barstar 的转基因油菜，这些植株便是 TA29－Barnase 基因工程不育油菜的恢复系。TA29－Barstar 的转基因油菜雄性可育、花药绒毡层发育良好，无任何不良表现。以 TA29－Barnase 转基因油菜做母本与 TA29－Barstar 转基因油菜杂交，得到同时含有 TA29－Barnase 和 TA29－Barstar 两种基因的后代，恢复雄性可育。与未转化的正常油菜相比，这些植株花药的绒毡层发育良好，能正常开裂并产生大量有功能的花粉粒，其形态、结构及萌发和受精功能皆正常。

第五节　动物的遗传转化

在许多情况下，大肠杆菌细胞中表达的真核生物蛋白质没有活性。这是因为许多真核生物的蛋白质在翻译后须经修饰才能有活性，修饰的类型包括形成

二硫键、蛋白质糖基化、信号肽的切除和某些氨基酸的修饰等。因此，在真核生物细胞中直接表达这些蛋白就显得十分必要。

一、动物转基因的方法

1. 显微注射法　显微注射法又简称微注射法，它是创造转基因动物的有效途径。其基本原理是用玻璃吸管固定受精卵，在显微镜下用特殊的 DNA 注射器将带有目的基因的 DNA 片段注入受精卵的雄原核中。然后，将带有目的基因的受精卵移入代孕母体内，使受精卵发育成完整的个体。

2. 逆转录病毒法　这一方法常用于老鼠的转基因研究。首先将外源目的基因克隆到老鼠的一种逆转录病毒上，让其侵染老鼠的 8 细胞胚。将处理的 8 细胞胚转入母鼠体内，生出的幼鼠一般为转基因嵌合体，经过进一步的杂交筛选就可得到纯合的转基因幼鼠。这种方法转化效率虽然很高，但由于病毒载体的容量有限，只能转移小片段 DNA。而且，逆转录病毒载体须插入到染色体上，其结果常导致宿主正常基因发生插入失活。这是这类载体的不足之处。

二、转化体的鉴定和分析

转化体的鉴定可通过多种方法进行，如报告基因检测、DNA 分子鉴定、RNA 分子鉴定、蛋白质分子鉴定和目的基因性状鉴定等。

检测报告基因（如 GUS、NOS、CAT 等）是较为方便的鉴定手段，用以确定外源基因的存在和表达。另外，报告基因在一些基础研究中应用也很广泛，如研究启动子的功能、基因调控元件的效应、可诱导基因的表达等。DNA 分子鉴定包括点杂交、Southern 吸印杂交和 PCR 扩增等。RNA 分子鉴定一般采用点杂交或 Northern 杂交，其操作步骤和原理都与 DNA 分子鉴定类似。DNA 和 RNA 分子鉴定依赖于碱基互补配对的原理，而蛋白质分子鉴定的基础是抗原抗体反应，主要有 Western 杂交和酶联免疫技术（ELISA）。如果是直接用目的基因转化，也可以只对目标性状进行检测。

三、重组 DNA 分子导入受体细胞

1. 昆虫细胞表达系统　在利用酵母表达系统时，仍有许多基因在酵母细胞中不能正常表达。后来，人们发展了昆虫和哺乳动物表达系统。

在昆虫表达系统中，最常用的是杆状病毒载体。杆状病毒能够侵染包括昆

虫在内的许多无脊椎动物。在侵染昆虫细胞后，杆状病毒既可以单个病毒粒子的形式释放，也可多个病毒颗粒一起由多角体蛋白包裹形成多面体。杆状病毒的单个病毒颗粒和多面体均具有侵染活性。多角体蛋白是由病毒基因编码的，它对于病毒的生活周期并不是必需的。在杆状病毒载体改造时，人们可以用外源目的基因替代多角体蛋白基因。

在杆状病毒侵染昆虫细胞 45 天后，病毒在宿主细胞中大量繁殖，外源目的蛋白也大量合成。在昆虫细胞中表达的高等真核生物蛋白产物往往能够得到正确的修饰，具有活性。目前，利用家蚕核型多角体病毒已在家蚕细胞中成功地表达了人 α－干扰素、白细胞介素、乙肝病毒表面抗原等多种外源蛋白。

2. *哺乳动物细胞表达系统*　脱分化的哺乳动物细胞也可以在离体条件下培养。通过特定的载体可以将外源基因转入培养的哺乳动物细胞系。在哺乳动物细胞的转化过程中常用的载体有猿猴病毒 SV40（图 9－36）、腺病毒、牛痘病毒等。将外源基因转入哺乳动物细胞常用的方法有磷酸钙沉淀法、电击法和脂质体介导法等。

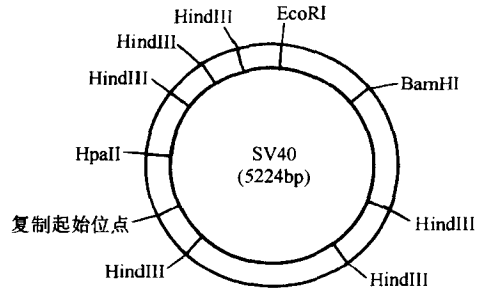


图 9－36　SV40 载体

四、动物基因工程的发展前景

应用基因工程技术大规模、低成本地生产动物生长激素具有巨大的经济潜力。现在科学家已经能够在大肠杆菌中生产牛生长激素（BGH）。若每天给乳牛注射这种激素，可使牛奶产量提高 40%，而且牛奶的质量没有变化。然而每天注射 BGH，显然十分耗费劳力。

胚胎显微注射技术最初是在培育转基因小鼠工作中发展起来的，这项新技术现在已被用于转基因家畜的研究。利用胚胎显微注射技术转移生长激素基因，有可能加快家畜的生长速度。随着分子生物学基础研究和 DNA 重组技术的发展，或许将来也有可能培育出改变了肉类品质的转基因家畜。例如现已发现，将牛的生长激素注射到乳牛或羊羔体内，可以改善这些转基因家畜的食物转换效率。提高蛋白质对脂类的比例，从而生产出更加符合人类要求的较瘦的肉类。再有，通过增加促性腺激素（gonadotrophin）的水平，或是将一种特殊的羊基因即能够控制产卵率或促卵率的促卵素（booroolla）基因导入动物体

内，就有可能培育出具有高产卵率特性的转基因家畜。这种促卵素基因是从澳大利亚绵羊中分离出来的，它能够提高绵羊的双胞胎和三胞胎的发生率。把这个基因导入其他牛和羊基因组中，可提高母牛和母羊的产仔率。此外，通过转基因技术引入角蛋白（kera－tin）基因，也可能使羊毛和皮革的质量得到改良。

然而要使用基因工程技术对动物的生长特性作出重大的改良，看来还需要好几年的努力才能有所突破。因为在作出上述发展之前，我们必须对动物的正常生理现象有更加广泛，更加深入的了解。在正常的情况下，动物体内的激素水平最终是保持平衡的，因此，按照人类的主观要求人为地改变某种激素的含量水平，破坏这种平衡，很可能会带来一系列麻烦。某些激素的内源含量水平被改变之后，可能也会影响到其他激素的水平。例如，已经发现，高水平的生长激素会使某些转基因小鼠的生育能力下降就是一个明显的例证。转基因动物的一种更为直接利用是，我们有可能将它当作是一种“生物工厂”（bio－factory），大量生产医用蛋白。例如人类血友病B是由于凝血因子IX基因突变引起的。传统的治疗方法是给患者输入血浆或纯化的凝血因子IX。但是这些方法都可能使患者再被染上其他的疾病，甚至是爱滋病，因此使用克隆来源的蛋白将具有很多的优势。研究者们现已将人类的凝血因子IX的基因与编码羊奶中的一种β－乳球蛋白（β－lactoglobulin，BLG）的控制序列连接，随后将这种重组基因注射到羊的受精卵中，再把受精卵植入母羊体内。果真如预期的相同，在该母羊产的小羊成熟以后，它们生产的羊奶里含有人类的凝血因子IX的成分。并且随后运用该技术还生产出来α1抗胰蛋白酶，从转基因羊的羊奶里得到了35克/升的酶蛋白。现在美国马萨诸塞州的金泽姆转基因公司已开始开发第一家制药农场，用转基因山羊生产人体蛋白。普遍认为，用转基因动物生产重组蛋白的费用将比传统生产方法低得多，而且设备投资也较小。预计本世纪人类将会使用在畜牧场生产的医药。现在通常采用动物血清提取生长激素和抗体，但牛奶的收集显然要比血清的收集容易得多。因此从上述转基因牛或羊的奶中提取生长激素和抗体具有极大的吸引力。

我国目前已获得转生长激素基因的转基因猪，其生长水平比非转基因猪提高20%。转基因鲤鱼已繁殖到第5代。据初步测定，这种转基因鱼群的生长比对照鱼群快20%，节约饲料10%。

五、动物体细胞克隆技术

1997年英国罗斯林研究所I.Wilmut等克隆出绵羊多利以来，克隆一词变的家喻户晓。高等动物是通过两性交配、精卵细胞结合而繁殖的。所谓克隆就

是指无性繁殖技术，也就是由一个体细胞的核经过不断分裂增殖长成一个新的动物个体的过程。除此之外，通过胚胎核移植、胚胎分割、胚胎嵌合等技术得到的动物也称为克隆动物。在绵羊多利的克隆中使用的是体细胞核移植的克隆方法。该过程主要是将体细胞经抑制分化培养，使其处于静息状态，采取体细胞核移植法，将其核导入除去染色体的成熟的卵母细胞中，将该卵细胞移入受体，经妊娠克隆出动物。理论上讲，这种方法可以无限制地克隆动物个体，因为分化了的体细胞核是取之不尽的。因此多利的降生才会震惊世界。这一技术可以培育大量的优良家畜，为人类提供某些药物和无免疫排斥的移植器官，并为医学实践提供了合适的动物模型。多利的遗传物质源于抑制妊娠绵羊的乳腺细胞，其中90%以上是乳腺上皮细胞，也有一些分化了的其他类型的细胞和少量没有分化的干细胞。多利的形态特征和供核羊相同，而与提供卵母细胞的母羊不同。它冲破了用胚胎细胞克隆动物的传统方式，有力地验证了细胞全能性理论，是克隆技术上的一次飞跃。

在2000年3月14日，曾参加克隆多利羊的英国PPL公司宣布他们利用克隆羊的技术又成功地克隆了5只猪。虽然克隆技术目前还不成熟，但预计在21世纪初这一技术将在畜牧生产上广泛应用。可以设想，利用克隆技术繁殖的一群基因型完全相同的高产畜禽将会大大提高畜牧业的生产率，为畜牧业带来一场真正的革命。特别是如果将动物克隆技术与转基因技术结合起来，人们就可以在短期内创造大量基因型相同的转基因动物，生产具有特定用途的目标产品。

第六节　重组DNA技术在医学上的应用

利用天然物质治疗人类疾病或达到某种医学效果一直是医学上的一个重要领域。但是生物体内具有重要功能的多种细胞素、激素、淋巴因子等的含量都很低，大量制备几乎不可能。自从70年代末产生了DNA重组技术以来，生物治疗才获得了飞跃发展。通过该技术可以产生大量的医学上急需的、而过去又难以生产分离的人体蛋白质，以用于疾病治疗。根据现阶段临床应用，生物治疗可以包括如下内容：细胞素治疗、抗细胞素治疗、免役保护治疗、毒素导向治疗、基因转录因子作为药物治疗、单克隆抗体治疗、寡核苷酸药物治疗、基因治疗及基因疫苗等9个方面。

一、基因治疗

基因治疗（gene therapy）是指将正常的外源基因导入生物体以达到治病

的目的。起初这一技术被用于治疗单基因缺陷的遗传病，如 1990 年开始用这种方法治疗先天性免疫缺陷症。现在已经有很多的实验室从事针对人类遗传疾病的克隆基因的研究，以期望通过植入正常的基因来克服遗传病。目前，基因治疗的对象已扩展到恶性肿瘤、艾滋病、乙型肝炎、心血管疾病和代谢性疾病等。美国联邦医疗卫生机构目前已批准了 100 多个基因治疗方案进入临床研究，其中 70% 的治疗对象为恶性肿瘤。由于直接将外源基因导入机体后的表达产物可产生良好的免疫反应，从而引申出基因疫苗的概念。

弄清基因组信息是进行基因治疗的前提条件，人类基因组是十分复杂的结构，单倍体有 30 亿个碱基对组成，大约有 5～10 万个基因，目前已鉴定出 1600 多个，其中已有 100 多个作了染色体定位。多数遗传病由此被弄清其遗传机理。1986 年由美达尔培哥提出癌症研究要弄清人类的基因组，国会批准了生物学的“月球探测器”计划即人体基因组计划（Human Genome Project）。我国也把它列入“863”高科技计划的一部分，已经把 10 万个基因的 30 亿个碱基分割成 50～100 个片段群，建立了 YAC 基因文库。不久前，美国已绘制出了 Y 染色体和 21 号染色体图谱。后者和老年痴呆症及唐氏综合症有关。

目前已知在全世界大约有 2000 种以上的人类疾病是由于单基因缺陷引起的，基因工程的发展，为治疗这些遗传疾病开辟了一条充满希望的新途径。基因治疗的基本构想是，应用 DNA 重组和基因转移等技术，把正常基因导入患者的体细胞中，合成出正常的基因产物来补偿异常基因的功能缺陷，从而使人类的遗传病得到纠正。

基因治疗中最合适的受体细胞是能够自我更新的造血干细胞群。这种细胞可容易地从骨髓中取出，在体外进行操作之后，再重新放回患者体内。例如镰形细胞贫血症，该病人体内的血红蛋白是异常的，即患者的 β－珠蛋白基因有缺陷。这就意味着用遗传工程的方法治疗该病将是可行的。因为血细胞是由骨髓产生的，我们可以通过用含有正常血红蛋白基因的骨髓来取代病人的异常骨髓。如果该过程使用病人自身的骨髓则是最好的，这样就解决了排斥的问题。但是我们如何又能够使正常的基因进入到具有异常基因的骨髓中去呢？克隆基因就是一个关键性的问题。把克隆的人类珠蛋白基因引入细胞中去的过程与把重组 DNA 加入到细菌细胞中一样。该实验主要过程包括以下几个过程：1）取出患者的部分骨髓细胞。2）在抽提的细胞中加入正常的 β－珠蛋白基因。3）培养细胞。4）用高剂量的 X－射线照射病人，杀死病人体内残存的骨髓细胞。5）尽快给患者植入转基因细胞，使其生长以代替原来骨髓的功能。因为没有骨髓细胞的病人将很快死亡。6）按照正常情况，正常基因的功能将足以克服患者的疾病。该计划被称为两步法，因为患者的细胞在体外被处理，然后

再将处理过的细胞送回体内。在波士顿一个患有白血病和镰形细胞贫血病的8岁女孩，医生通过杀死其癌变的骨髓细胞并植入他人的正常骨髓细胞，结果不但治好了她的白血病，同时也使她的镰形贫血症得到了治疗。

除了镰形细胞贫血症以外，还有数种遗传病可以使用组织转移的方法进行治疗，因此基因治疗法的临床应用是能够成功的。另一个引起人们注意的是由于缺乏腺苷脱氨酶（adenosine deaminase，ADA）而引起的一种极为罕见的SCID先天免疫缺陷综合症。ADA缺陷最主要是影响造血细胞，使之积累脱氧核苷导致未成熟的T细胞解体。在正常细胞中，ADA酶是组成型表达，而且只需要少量即可阻止dATP的积累。这说明用基因治疗术治疗SCID免疫缺陷综合症需要发展出一种能够确保转移的ADA酶基因处于严格表达控制的载体分子上。1995年，法国的Anderson成功地给一名患者进行了基因治疗，使患者存活五年以上。

基因治疗的另一个目标是莱纳二氏综合症（Lesch－Nyhan syndrom），这是一种极为少见的自毁容貌的遗传病，它是由于次黄嘌呤、鸟嘌呤、磷酸核糖基转移酶（HGPRT）缺陷造成的。这种酶存在于造血细胞及中枢神经系统中，因此可以通过基因工程的方法改变骨髓细胞，使之表达HGPRT，但对这种病人的中枢神经系统的缺陷则很难纠正。

目前人们还试图运用转基因技术来治疗癌症，试图通过外源基因来增强患者自身的免疫系统的对癌细胞的杀伤力。但这首先需要分离那些能够渗透入肿瘤中、并可能杀死肿瘤的白细胞，还须把肿瘤坏死因子的基因转入该白细胞，该基因的蛋白产物只能杀死肿瘤细胞而不伤害正常组织的细胞。

二、爱滋病的治疗

获得性免疫缺陷综合症（acquired immune deficiency syndrome）简称爱滋病（AIDS），为人类的一种致命性传染疾病，自1980年在美国洛杉矶发现以来，正在以惊人的速度向全球传播。AIDS最早于20世纪50年代中期在非洲出现，最初的感染源可能是猿猴。流行病学家认为爱滋病由非洲传到加勒比海地区，然后再传到美国，自此以后，又传到加拿大、欧洲大陆及亚洲地区。迄今为止还没有抗HIV病毒的有效疫苗，也没有找到能够完全抑制此种病毒的药物。目前已采用了多种方法，包括基因操作和转基因技术，研究HIV病毒的分子生物学及病理原因，以期早日生产出疫苗和治疗药物。

已如第六章所述，HIV－I病毒有3个大的编码区，即gag（核心抗原）、pol（复制酶及整合酶）和env（外膜糖蛋白）。此外，还有若干个其他较小的

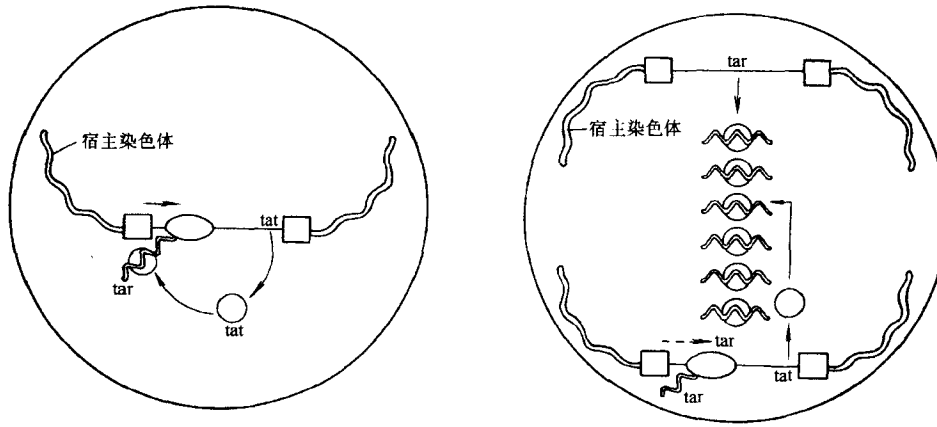
可读框，它们所编码的蛋白质参与控制 HIV 基因的表达活性。tat 基因编码一种较小的、86 个氨基酸的蛋白质。这个蛋白质可以加强病毒基因的表达。HIV 的 5′－LTR 之 R 区段内的序列叫做 tar 的结构参与控制 tat 基因的表达。trs 基因与 tat 基因共用一条 mRNA，但按不同的编码结构进行表达，所以 trs 编码的是一种与 tat 完全不同的有 116 个氨基酸的蛋白质，它同样可以增强病毒基因表达与复制的能力。

预防 HIV 感染需要疫苗，目前正在进行有关爱滋病疫苗的研制。第一步是鉴定能够诱导产生保护性抗体和细胞免疫的潜在抗原。HIV 蛋白已在许多细胞系种实现了表达，例如 Env、Tat 和 Sor 蛋白已能在大肠杆菌细胞中表达合成；Gag 和 Pol 蛋白已在酵母细胞中表达合成；Env 蛋白已在痘苗病毒载体和酵母中表达。对所有这些由重组 DNA 产生的抗原都已经作了检测，以查清它们是否具有诱导中和抗体的能力。生产疫苗的重组 DNA 技术，对绝大多数的病毒来说都是相同的。然而发展 Env 蛋白疫苗的主要问题在于一个病毒株的分离体和抗体之间有很大的差异，而且并不总能抗其他病毒株的侵染。

发展控制或治疗爱滋病的药剂的基本原理是，从病毒的生活史中鉴定出一个独特的阶段，在此时使用药物治疗剂既可抑制病毒的生长，而对正常细胞的代谢活动又没有严重副作用的药物。爱滋病毒生命周期中，有某些方面是这种病毒所特有的，因而是使用药剂治疗的好机会。此外，由于反转录酶在酵母细胞中具有正常活性，所以用这个系统可以发展出快速的测定抑制作用技术。基因转移技术在爱滋病预防与治疗中的另一途径是研究控制 HIV－1 病毒基因表达的各种因子。因为只要 tat 基因产物及其活性被抑制，病毒的复制也就完全或大部分被抑制。

现在科学家试图通过把 HIV 病毒作为一种“特洛伊木马”，把抗性基因转入 HIV，并通过该病毒侵染细胞的方法来治疗爱滋病。采用这一策略首先需要构建 HIV 载体，如图 9－37 所述的方法。将病毒中的 gag，pol 和 env 切除，只留下其 LTR 序列，然后再把人类的腺苷酸脱氨酶（adenosine deaminase，AD）基因插入病毒载体中。经过改造的病毒对正常细胞没有伤害，因为它缺乏 HIV 复制所必须的几个基因。下一步还须将一个抑制病毒基因表达的基因置入 HIV 病毒载体中，如 tar 基因，它编码一段与 tat 蛋白结合的 RNA 序列。用这种 HIV 载体来转染 HIV 病毒侵染的细胞，那么，它所产生的大量的 tar RNA 序列就会吸引 tat 蛋白并与之结合，阻止 tat 蛋白行使其正常的启动 HIV 复制的功能，从而达到抑制病毒繁殖的作用（如图 9－37）。

另一个设想是采用反义 RNA 来阻断爱滋病毒的复制。有许多研究都已表明，将一种待阻断基因的反义 RNA 导入细胞，便可阻断该基因的表达。很显



(a)正常的HIV感染　　(b)"特码"策略

图 9－37　用 HIV 载体治疗爱滋病的策略

A. 在正常的 HIV 侵染中，tat 基因产生的 tat 蛋白与病毒转录子的 tar 区结合，刺激转录进行　B. "特洛伊木马"方法是构建只含有 tar 序列的逆转录载体，产生的 tar RNA 可以吸引由正常病毒产生的 tat 蛋白，而与病毒转录子的 tar 区结合的 TAT 蛋白质就很少，从而使转录变慢或抑制病毒的复制

然这是由于形成了一种特殊的双链 RNA－RNA 分子，而使翻译受阻的缘故。构建一种非致病的 T 细胞病毒载体，就能够将反义 RNA 基因导入细胞，从而有效地干扰 HIV 病毒的复制。

三、生物技术新药

应用重组 DNA 技术制成的第一种药品是治疗糖尿病的人胰岛素，现已得到广泛的应用。干扰素、血栓溶解剂 TPA、尿激酶、白细胞介素和一些单克隆抗体等生物技术新药也相继上市。新型的诊断试剂盒也是生物技术新药的一个发展方向。目前已开发了多种用于临床的诊断试剂盒，可以用于各种肿瘤的早期诊断、妊娠和各种重要疾病的诊断。

如第六章所述，干扰素是一类重要的细胞素，是病毒和细胞间拮抗作用（又称干扰现象）的主要介质，它在同种细胞上具有广谱抗病毒、抗细胞分裂及免疫调节等多种生物学功能。经 10 多年的临床应用研究表明，干扰素活性的发挥受细胞基因组的调控，与机体内许多 RNA 及蛋白质的合成有关。研究证实，人、小鼠、牛、马等哺乳类的干扰素均有 α、β、γ、ω4 个类型。其中 α 型又可分为两大家族和许多亚族，人 α 干扰素的第 I 类家族（HuIFN－αI）有 15～20 个基因组成，其中大部分都编码功能蛋白，在核苷酸水平上有 90% 的

同源性；而人 α 干扰素的第 II 类家族（HuIFN－αII）有 5～6 个成员，与 αI 家族有 50％的同源性，其中大部分是假基因；人 β 干扰素（HuIFN－β）和 γ 干扰素（HuIFN－γ）都只有一个亚型。1986 年，美国食品药品管理委员会首先批准由基因 αIIa 和 αIIb 编码的干扰素投放市场，基因工程 β、γ 干扰素也相继于 1990 年、1993 年获准投放市场。目前已有 57 个国家批准干扰素上市治疗约 30 多种疾病。我国研制开发了的 αIb 型、αIIa 型基因工程干扰素也已经国家新药审评委员会审议通过，投放市场，它是我国第一批经国家正式批准的基因工程高技术药物。目前我国正式批准上市的基因工程药物见表 9－8。

表 9－8　我国正式批准上市的基因工程药物

名　　称	生产时间	治疗疾病
RhuIFNa1b（外用）	1989 年试生产	病毒性角膜炎
RhuIFNa1b	96 年正式生产	HBV、HCV 等
RhuIFNa2a	97 年试生产	疱疹、HBV、HC 等
RhuIFNr	94 年，95 年试生产	类风湿
RhuIL－2	97 年正式生产	癌症辅助治疗
RhuGCSF	95 年试生产	化疗生白细胞
RhuGMCSF	97 年试生产	化疗生白细胞
Rsk（链激酶）	96 年试生产	心肌梗塞溶栓
Rhu EPO	97 年试生产	再生障碍性贫血
碱性成纤维生长因子 bFGF	96 年试生产	创伤、烧伤的治疗

四、基因工程与疫苗生产

有许多传染性疾病如乙型肝炎、疟疾、非洲淋巴癌、狂犬病以及爱滋病等严重威胁到人类的生命安全。还有许多疾病虽然并不致命，但也会给社会和经济发展造成不良的影响，例如生殖疱疹病毒以及流感病毒的感染就是属于这方面的两个明显的例子。此外一些家畜的重要的传染病如口蹄疫等也会给生产和生活带来很大损失。对于这些疾病，目前尚无有效治疗手段，最好的办法是进行免疫注射，预防初期感染。

近年来用基因工程技术生产的一批新疫苗已得到广泛应用。它可以代替花费昂贵而有时还相当危险的传统工艺，生产出能够抵抗这些疾病的疫苗。如我国生产的基因工程乙肝疫苗于 1992 年获国家批准投放市场，目前已在城镇新生儿全部实行计划免疫。基因工程乙肝疫苗的推广应用已使我国城市婴儿的乙

肝带毒率大幅度降低，为控制乙肝病毒的流行发挥了重要的作用。目前还有许多人、畜疫苗已进入临床试验阶段，预计在不远的将来基因工程疫苗将得到更为广泛的应用。

1. 重组亚基疫苗　所谓亚基疫苗是指它只含有病原体的一个或几个抗原，而不含其他遗传信息。从原理上讲，用这些疫苗注射动物，都可使之获得抗性，免受病原体的感染。亚基疫苗的优点是不含感染性组分，因而无须灭活，而且也无致病性。一些细菌性疾病的亚基疫苗如破伤风、霍乱、伤寒的类毒素疫苗等都已证实是可行的。

制备重组亚基疫苗的一般方法是应用基因工程技术，克隆特异的中和抗原的编码基因，并使之在适当的寄主－载体系统中表达。所产生的抗原务必经过纯化，但不必作无毒测试，因为按这种方法生产，不可能有活病原的污染。

目前世界各地有大量的人口感染了乙型肝炎病毒。在东亚和热带非洲至少有10%的人受到HBV的影响，而北美每年都要增加20万人的乙型肝炎患者。此外，HBV感染通常还与特殊的肝癌有着密切的关系。据统计每年全世界死于肝癌的病人有30万人左右。肝炎病毒具有高度的寄主专一性，只能感染人类和黑猩猩。这意味着我们唯有从肝炎患者身上才能获得有限数量的病毒供作疫苗使用。

现在，乙肝疫苗的预防效果与安全性方面都已经在乙肝高危人群和新生儿中得到了证实。既然如此，那为什么还要发展其他类型的抗HBV的新疫苗呢？其原因主要是HBV不能在组织培养中增殖，只能从患者血液中提取，因此来源受限制；其次，人血的收集十分困难，而且还需要复杂的提纯和灭活过程，最后还要在黑猩猩身上作安全注射检验，这样制备的抗HBV的疫苗的价格就十分昂贵了。此外，由于爱滋病的威胁，许多人担心使用血浆源性的疫苗会人为地促进爱滋病的传播。因此迫切需要生产出一种替代的新疫苗，这样人们便自然而然地把希望寄托在重组DNA疫苗的发展上。

乙肝表面抗原HbsAg是由S基因编码的。最早人们是把S基因克隆在大肠菌细胞中进行表达，虽然可以合成出一些蛋白质，但其产量水平很低，而且也不能组装成免疫原性的颗粒（immunogenic particle），因此后来便转向在酵母和动物细胞中生产HbsAg。

目前生产乙肝重组亚基疫苗的方法有三种：其一就是由酵母生产的，P.Valenzuela等人（1982）将S基因的一个835bp片段克隆在一种表达载体上，结果在酵母中合成出了HbsAg颗粒，其产量达25μg/L。应用新的高效表达载体，其产量至少可以提高2.5mg/L。由酵母生产的HbsAg颗粒经纯化之

后，注射到黑猩猩体内，可以使其对活的乙型肝炎病毒的感染产生完全的抗性。于是，在 1984 年一种利用基因工程技术生产的新疫苗便告诞生。酵母表达系统现在已经能够大规模生产供给人类使用的第一代重组肝炎疫苗。第二种生产方法是在动物细胞中进行。最有希望的是利用 dhfr 扩增载体在 dhfr－CHO 细胞中生产 HbsAg 颗粒。采用动物细胞系统生产 HbsAg 可能比酵母系统要优越些，一个主要的优点是 HbsAg 颗粒可以直接分泌在培养基中，这样纯化就比较简单，生产过程是连续的，而且动物细胞分泌的 HbsAg 还具有更强的免疫原特性。由动物细胞生产重组肝炎疫苗目前已处在临床试验阶段。第三种生产方式是应用一种能在人体细胞内复制的病毒载体，建立活疫苗。例如有人已应用一种痘苗病毒作载体克隆 S 基因，建立一种活疫苗。感染了这种痘苗病毒重组体的细胞能够合成和分泌 HbsAg，感染的兔子则产生抗 HbsAg 的抗体。感染重组体疫苗的黑猩猩不产生抗 HbsAg 反应，但可以防止 HBV 引起的乙型肝炎。

上面所述的生产 HbsAg 的方法同样也可以用来生产任何一种病毒的抗原。最近不少发达国家已投入了很大的力量来发展抗 A 型流感病毒的亚基疫苗。这种病毒引发众所周知的呼吸道疾病，即流行性感冒。这种疾病有很高的发病率，严重的可以致死，对人类的健康是一种很大的威胁。然而至今还没有一种真正有效的抗流感病毒的疫苗。

发展抗流感病毒的疫苗有许多困难，最主要的是这一病毒本身存在着极大的遗传变异性。它有两种表面抗原，即血细胞凝集素（HA）和神经氨酸苷酶（N）。HA 的抗体对免疫性最为重要，但如果同时也产生 N 抗体，那么免疫性保护作用则会得到极大的提高。至今已经鉴定出了 13 个不同的 HA 基因。每个都编码一种独特的抗原蛋白，叫做亚型抗原蛋白。作用于一种 HA 亚型抗原的抗体并不能作用于别的 HA 亚型抗原。此外还有 9 种 N 亚型抗原。具有任何一种 HA 和 N 型组合的流感病毒都可以出现，所以病毒的类型特别多。病毒的一种特殊亚型可以在数年内反复出现，然后突然被一种全新的亚型所取代。这种现象叫做抗原漂移或抗原更换（antigenic shift）。人群对这种新亚型的流感病毒不具免疫性，于是上年制备的疫苗便完全无用了。此外，由于 HA 基因突变的结果，病毒抗原也会逐渐发生变化，产生出不能被抗体有效中和的抗原，这叫做抗原漂变或抗原漂流（antigenic drift）。现在已经发展出好几种用来表达流感病毒表面抗原的寄主—载体系统。

2. *重组病毒活疫苗*　接种痘苗病毒防治天花是使用非重组病毒活疫苗的一个最典型的例子。近 200 年来，它一直被用来保护人类免生天花。然而在 1979 年 10 月 26 日世界卫生组织（WHO）宣布，天花已经从世界上根除，因

此继续接种痘苗疫苗已无必要。现在的目标是应用重组 DNA 技术扩大痘苗病毒潜力，使之成为能够表达其他病毒抗原的真核基因载体。基本方法是将病原体中和抗原基因的编码序列插入到痘苗病毒的基因组上，形成一种能够在人体细胞中复制，同时又能表达痘苗抗原及其他抗原的重组病毒。这种体系具有活疫苗的全部优点，而又不存在使用病原体所可能出现的危险。

应用痘苗病毒系统，目前已经发展出了许多种抵抗重要病毒疾病的潜在疫苗，其中包括狂犬病毒、Ⅰ型及Ⅱ型疱疹病毒、流感病毒、乙肝病毒、EB 病毒、水泡性口膜炎病毒（VSV）以及爱滋病毒等。同样也可能生产出多价重组痘苗病毒，例如现在已经成功地把 HSV－lgD、HbsAg 及流感病毒 HA 基因同时插入在同一个痘苗的基因组上。接种了这种多价重组痘苗病毒的兔子所产生的抗体能够同这 3 种抗原发生作用。到目前为止，在痘苗病毒载体上表达外源抗原的成功率是 100％。

五、重组 DNA 探针与遗传病诊断

重组 DNA 技术对临床医学的另一个重要贡献是为疾病的诊断特别是人类遗传性疾病的诊断提供了一种简便快速、特异性强、灵敏度高的 DNA 探针检测法。所谓 DNA 探针是指用同位素，荧光分子以及化学发光催化剂等标记的单链 DNA 片段，它可以同被检测的 DNA 分子中的同源互补序列杂交，从而检出所要查明的 DNA 或基因。

在人类已经认识的遗传性疾病中，唯有遗传性血红蛋白机能障碍症已经在 DNA 水平上作了广泛深入的研究。最突出的一个例子是，发现β－珠蛋白链的第 6 个氨基酸残基的密码子发生了由 GAG 到 GTG 的突变。在纯合子中，这种突变的结果便产生出镰形细胞贫血症。在临床医学上最重要的血红蛋白病是地中海贫血症和镰形细胞贫血症以及它的许多变型。

地中海贫血症是由于珠蛋白链合成不平衡（它们当中有一条链合成速度下降或不合成）而引起的一类遗传性疾病。许多α－地中海贫血症状是基因缺陷造成的（当然也有少数几个非缺失的例子）。β－地中海贫血症比较复杂，已经鉴定出来的有 20 多种不同的分子损伤，其中大多数是在β－珠蛋白基因的可读框发生了无义突变或移码突变，其他则是影响转录加工的点突变，或是影响启动子的点突变，仅有一例是由于β－珠蛋白基因缺失所致。

对于上述这些疾病及其他许多的遗传疾病，目前都还没有定形的治疗方法。通常只能在一个病变的婴儿出生之后，才能断定其双亲是遗传性疾病的携带者。当我们根据已出生的孩子或其他因素确定其孕妇怀有病变胎儿危险性

时，便可以采用 DNA 探针检测法进行产前诊断，从而确定是否中止妊娠。

下面我们以血红蛋白病为例说明 DNA 探针在遗传病诊断中的应用。一般认为用重组 DNA 技术对此种疾病作产前诊断可作为其他遗传病诊断的一个范例。当前血红蛋白病产前诊断程序是，首先进行胎儿血液取样，然后通过放射性标记估算珠蛋白链合成的相对速率。此法的缺点是胎儿血液取样需待怀孕 18 周后才能进行。通过羊膜穿刺法也可得到胎儿细胞，围绕胎儿的羊水中含有胎儿的尿和其他分泌物，大约在怀孕的 16 周取出羊水样品。胎儿细胞作体外培养，以进行生化和染色体分析。除血红蛋白病变外，这些分析还能用于鉴定其他病变，如某些酶的缺陷或唐氏综合征(Down's syndrome)。胎儿细胞还可用来制备 DNA 供作限制性片段长度多态性分析。

很明显，如果一个突变会导致基因组 DNA 上减少或增加一个限制酶的识别位点，它就可作为突变与否的一种标记。在镰形细胞贫血症中，从 GAG 到 GTG 突变，消除了 Ddel 或 Mstl 2 种限制酶识别位点。因此，用限制酶 Ddel 和 Mstl 分别消化突变的和正常的 DNA，然后再用克隆的 β－珠蛋白 DNA 探针进行 Southern 杂交，便可以检测出 GAG－GTG 的突变。此法只适用于某些改变了一个限制位点，或因一个大的缺失或因重排而改变了限制酶限制图谱的遗传性疾病检测，但不适用于大多数 β－地中海贫血症的检测。

在整个 β－珠蛋白基因族中，散布有许许多多限制性位点，这些结论是根据由 Southern 杂交实验所揭示的限制片段长度多态性现象得出的。这些特性可以作为产前诊断的连锁记号，也就是说，与 β－珠蛋白基因在形体上紧密连锁的多态位点是能够伴随该基因一道遗传的。这些多态性可以按如下两种方式加以利用。第一，有些多态性是与特定的珠蛋白突变连锁的。例如在美国黑人群体的镰形突变中大约有 60％是与一个紧挨于珠蛋白基因（它已失去了 Hpal 位点）的多态性位点突变连锁。这种连锁可以作为该遗传病诊断的基础，但不适于应用 Mstl 进行直接分析。这种连锁形式即所谓的连锁不平衡（linkage disequilibrium）是很少见的。第二，在产前诊断之前，须进行家族研究以确定多态限制位点和一个特定的 β－珠蛋白基因突变间的连锁关系。但这种方式并不总是可行的，因为也许就不存在合适的多态记号。据报道，在英国这种方式只适用于不到 50％的 β－地中海贫血症病例的产前诊断。

最近已发展出了一种强有力的直接分析点突变的方法。具体做法是，合成出两个 19bp 的寡核苷酸，一个与正常 β－珠蛋白基因（βA）的氨基末端区互补，另一个与镰形细胞 β－珠蛋白基因（βS）互补。这些寡核苷酸经同位素标记后用作 Southern 杂交的探针。在合适的条件下，这些探针能区分出正常的和突变的等位基因。正常纯正合子的 DNA 仅与 βA 探针杂交；镰形细胞纯合

子的 DNA 仅与 βS 探针杂交；而唯有杂合子的 DNA 能够同时与两种探针杂交。因此，这些实验表明，应用寡核苷酸探针，便能够将一条完全互补的 DNA 同一条仅含一个错配碱基的 DNA 区分开来。这一方法具有普遍的适用性，只要能够确定突变位点（可以是置换、插入或缺失）附近的核苷酸序列，此法就可适用于任何遗传病的诊断。

小　结

重组 DNA 技术就是通过特殊的酶，将不同来源的 DNA 片段在体外进行重组，产生新的重组 DNA 分子，并将这些重组 DNA 分子转入适合的生物受体后可以扩增或表达。

该技术通常是从供体组织抽提并获得适合的大 DNA 片段，将外源 DNA 片段与载体用同一种限制性内切酶处理，再用连接酶将外源的 DNA 分子插入到能够自我复制的载体 DNA 分子中，构建成重组 DNA 分子，随后将重组 DNA 分子导入宿主细胞，通过各种方法选择携带理想重组 DNA 分子的受体细胞。

限制性核酸内切酶是进行 DNA 分子重组的重要的工具酶类，能够识别 DNA 分子特定序列，并在这一序列或序列附近将 DNA 分子进行切割的内切核酸酶。从病毒到人类的 DNA 分子都随机包含着一系列限制性内切酶的作用位点，因此可以根据酶切位点可以将大分子的 DNA 剪切成适当大小的片段。将这些片段插入到克隆载体上，导入大肠杆菌中进行克隆，就可以构建出该物种的基因文库。所谓 cDNA 即利用 mRNA 为模板和逆转录病毒的逆转录酶而复制的一种互补 DNA。将 cDNA 用适当的限制性内切酶处理，然后将 cDNA 片段插入到适当的载体中进行克隆，就可以得到 cDNA 文库。在基因克隆过程中，已有很多方法可用来选择目的基因。利用 cDNA 为探针，与细菌菌落杂交是鉴定和分离目标克隆的重要方法之一。通过功能性互补、定位克隆、标签法等也可以寻找目的基因。

作为载体，一个最基本的要求就是具有一个复制起始位点，能够在寄主细胞中进行自我复制，这样可使目的基因得到增殖，并传递到后代中。目前常用的载体有质粒、噬菌体和黏粒等。质粒是一种双链环状 DNA 分子，它带有一个复制起始位点，能够在细菌中独立进行自我复制。单链噬菌体克隆载体主要有大肠杆菌的丝状噬菌体 M13、fl 和 fd，还有对噬菌体进行改造得到的 M13mp 系列载体。双链噬菌体载体以 λ 噬菌体为主。黏粒载体是一种含有 λ 噬菌体 cos 区域的质粒，这种载体的设计是为了克隆和繁殖大片段的染色体

DNA。酵母和细菌人工染色体是最新一代的载体系统，在克隆大片段 DNA 上是前者所不能取代的。

DNA 核苷酸序列分析法为我们了解基因结构与功能以及基因克隆的一重要方法。目前较常用的主要为 Sanger 双脱氧法。

DNA 聚合酶链式反应使我们能够从复杂的 DNA 分子群体中有选择地复制一段特异序列，应用范围广，主要应用于基因组 DNA 或 mRNA 序列的直接测定、基因分离、克隆、基因突变的直接检测等等。

重组 DNA 技术结合动植物细胞和组织培养、酶的固定化、现代化生物反应器和计算机技术的应用以及产品分离、纯化等技术形成了现代生物技术。生物技术目前也已被广泛地用于医学、食品、化学、农业及环境保护等领域，为这些行业带来了一场新的技术革命，产生了生物技术这一高新技术产业。

将外源重组分子导入受体细胞的途径包括转化（或转染）、转导、显微注射和电穿孔等多种不同方式。转化和转导主要应用于细菌一类的原核细胞和酵母这样的低等真核细胞，而显微注射和电穿孔则主要应用于高等动植物。

利用基因工程技术可以把控制特定性状的基因转入现有植物品种，大大缩短育种周期，创造出用常规育种方法很难得到的新型优良品种，因此在农作物品种改良中展现了广阔的应用前景。应用基因工程技术改变作物的某些遗传特性，使它们获得新的丰产优质的遗传性状是改良农作物生产的最直接的办法。

绿色植物光合作用的产物是人类社会的主要物质和能量的来源。如何有效地提高光合作用的效率已成为植物基因工程研究活跃的领域。现在许多科学家将 RuBisCo 酶作为基因工程的目标之一，以期通过提高 RuBisCo 同 CO_2 的亲和力来改善光合作用的效率。叶绿体的基因操作目前也引起了普遍的关注，人们设想通过改变光合系统 I 和光合系统 II 反应的中心蛋白质基因 psaA 和 psbA 来改变植物的光能转能效率，使其达到最佳的水平，从而提高光合作用的效率。

固氮根瘤菌可以在豆科植物根部形成根瘤，进行固氮作用。固氮作用是在根瘤中的固氮复合酶的作用下完成的。固氮复合酶可以将 N_2 还原为可被利用的 $NH^{+}4$。把非豆科植物特别是禾谷类农作物转变为固氮植物是植物固氮基因工程的一重要的目标。

禾谷类植物和豆科植物的种子含有丰富的贮存蛋白，是人类重要的蛋白质来源。科学家设想应用重组 DNA 技术改变种子贮存蛋白编码基因的特定碱基序列结构，使之加上缺少的人类必需氨基酸的密码子。

病毒感染使许多农作物的产量显著下降，给农业生产造成巨大的损失。理想的办法是应用基因工程技术，从自然界中鉴定出抗病毒的基因，然后通过重

组、转化等手段，培育出携带并能遗传抗病基因的转基因植物。

培育抗病虫害的抗性农作物新品种是一种既经济实惠，又无不良副作用的有效途径。应用基因操作技术，创造出具有抗病虫害的农作物新品种具有明显的经济效益。在抗虫作物育种方面一条可能的途径是，直接将苏去金杆菌毒素基因导入植物，使之获得抗害虫的遗传特性。

抗除草剂基因工程植物目前在国内外倍受重视。除草剂是一类用于清除大田杂草的化学药剂，但它往往也会抑制作物的生长。将外源抗除草剂基因导入植物基因组，作物本身就具备了抗除草剂的特性，除草剂的使用就会更加方便有效。

用植物基因工程的方法还可以获得雄性不育株系，以抗除草剂基因与不育基因紧密连锁，就可用除草剂消除不育群体中的可育件，繁殖高纯度不育系。

哺乳动物细胞可以在离体条件下培养，通过特定的载体可以将外源基因转入培养的哺乳动物细胞系。

应用基因工程技术大规模、低成本地生产动物生长激素具有巨大的经济潜力。在转基因家畜的研究中，利用胚胎显微注射技术转移生长激素基因，有可能加快家畜的生长速度。转基因动物的一种更为直接的利用是可以将它当作是一种“生物工厂”，大量生产医用蛋白。

利用天然物质治疗人类疾病或达到某种医学效果一直是医学上的一个重要领域。但是生物体内具有重要功能的多种细胞素、激素、淋巴因子等的含量都很低，通过重组 DNA 技术可以产生大量的医学上急需的、而过去又难以生产分离的人体蛋白质，用于疾病治疗。

基因治疗是指将正常的外源基因导入生物体以达到治病的目的。起初这一技术被用于治疗单基因缺陷的遗传病，目前，基因治疗的对象已扩展到恶性肿瘤、爱滋病、乙型肝炎、心血管疾病和代谢性疾病等。基因治疗的基本构想是，应用 DNA 重组和基因转移等技术，把正常的基因导入患者的体细胞中，合成出正常的基因产物来补偿有缺陷基因的功能，从而使人类的遗传病得到纠正。

获得性免疫缺陷综合症简称爱滋病（AIDS），已成为影响世界各国人民健康的一种严重的疾病。目前采用包括基因重组和转基因技术在内的多种生物学技术，研究 HIV 病毒的分子生物学及病理原因，以期早日生产出疫苗和治疗药物。

应用重组 DNA 技术制成的第一种药品是治疗糖尿病的人胰岛素，现已得到广泛的应用。干扰素、血栓溶解剂 TPA、尿激酶、白细胞介素和一些单克隆抗体等生物技术新药也相继上市。用基因工程技术生产的新疫苗已得到广泛

应用。

重组 DNA 技术对临床医学的另一个重要贡献是为疾病的诊断，特别是人类遗传性疾病的诊断提供了一种简便快速、特异性强、灵敏度高的 DNA 探针检测法。

复　习　题

1. 如何建立某种生物的基因文库？如果已知另一生物的一基因序列，如何从基因文库中寻找该基因克隆？

2. 何谓 cDNA 文库？cDNA 文库与基因文库有何差别？

3. 已知某蛋白质的氨基酸序列，如何寻找目的基因？

4. 试述 DNA 序列分析的主要方法的原理及步骤。

5. 试述 DNA 聚合酶链式反应的原理。

6. 重组 DNA 技术在农业上都有哪些应用？

7. 重组 DNA 技术在医学领域都有哪些应用？

中文索引

癌 carcinoma 283
癌病毒 oncovirus 281
安慰诱导物，义务诱导物 gratuitous inducer 133
氨酰基－tRNA 合成酶 amino acyl－tRNA synthetase 18，134
氨酰基位点 aninoacyl site 21
白细胞干扰素 leukocyte interferon 330
摆动假说 wobble hypothesis 23
半保留复制 semi－conservative replication 10
半染色单体交叉 half chromatid chiasma 69
半衰期 half life 194
棒病毒 rhabdovirus 277
胞嘧啶 cytocine，C 4
背腹极性系统 dorsal－ventral systerm 421
背腹轴 dorsoventral 417
被刻印基因 imprinted gene 214
鼻病毒 rhinovirus 276
闭合环状 DNA closed circular DNA 100
编码序列 coding sequense 18
变性 dissociation 13
标签法 tagging 468
表达文库 expression library 465
表达载体 expression vector 483
表面糖蛋白变体 variable surface glycoprotein 412
别乳糖 allo－lactose 132
丙酮酸激酶 pyruvate kinase 202
并列控制 coordinate control 51
病毒 virus 265
病毒癌基因 V－oncogene 316
病毒基因启动子 virus gene promotor 302
病毒基因转录终止 virus transcriptional termination 306
补给模型 recruitment model 118
补骨脂素 psoralen 31
不等交换 unequal crossing over 68
不均一核 RNA heterogeneous nuclear RNA，hnRNA 18
不相容性 imcompatibility 113
操纵元 operon 51
操作子 operator 52
侧翼序列 flanking sequense 10
插入 insertion 57
插入序列 insertiomal sequense 76
插入因子 insertional element 60
差别基因表达 differential gene expression 214
长末端重复 long terminal repeat，LTR 283

肠病毒 enterovirus 276
常染色质 euchromatin 148
超螺旋 supercoiling 8
沉默子 silencer 165
成熟促进因子 maturation promoting factor 434
成纤维细胞干扰素 fibroblast interferon 330
疮病毒 aphtho virus 276
次黄嘌呤 hypoxanthin 63
错义突变 missense mutation 58
大 T 抗原 large T antigen 287
大沟 major grove 7
单拷贝质粒 single copy plasmid 112
单克隆抗体 monoclonal antibody 348
单链 DNA 结合蛋白 single - stranded DNA binding protein, SSB 8
单链结合蛋白 single stranded binding protein 8
蛋白质释放因子 protein release factor, RF 27
倒位 inversion 88
颠倒重复 inverted repeat 54
颠倒重复序列 inverted repetitive sequence 14
颠换 transversion 57
点突变 point mutation 57
淀粉体 amyloplast 247
调节发育 regulative development 428
调节基因 regulatory gene 51
调节元 regulon 51
定位克隆 postitional cloning 467
动基体原生动物 kinetoplastid protozoa 236
动力竞争 dynamic competition 180
痘病毒 poxvirus 270
端粒 telomere 150
端粒基因 telometric gene 414
端粒基因 telometric gene 414
端粒酶 telomerase 156
多聚 A 核酸酶 poly A ribonuclease, PAN 195
多聚 A 聚合酶 polyA polymerase 183
多聚 A 尾巴 poly A tail 18
多拷贝质粒 multicopy plasmid 112
多瘤病毒 polyomavirus 270
多顺反子 mRNA polycistronic mRNA 18
二氢叶酸还原酶 dihydrofolate reductase 161
二型蛋白基因 diplomorphic gene 46
二元载体 binary vector 494
发育定时钟 developmental timer, DT 378
翻译 translation 25
翻译操作子 translational operator 126
翻译偶联 translation coupling 126
反基因组 antigenome 280
反密码子 anticodon 18, 22
反义 RNA antisense RNA 113
反终止子 antiterminator 120
反作用抑制子 trans - acting suppressor 83
反作用子 trans - acting element 52
泛素 ubiquitin 203
非孟德尔遗传 non - Mendelian inheritance 228
非自主因子 nonautonomous element 81
分解代谢活化蛋白 catabolite activator protein, CAP 130
分散排列 dispersed array 14
分支点序列 branch point sequense 184
辅助性 T 细胞 helper T lymphocyte, TH 296
负超螺旋 negative super coiling 8
附加体 episome 112
附属合成 bypass synthesis 38
复合基因 complex gene 48

复性 reassociation 13
复制叉 replication fork 10
复制启始区 replication origin11，152
复制启始区 replication origin 11，152
复制胜任型 replication competent 283
复制体 replisome 110
复制型 replicative form 274
复制型 RNA 链 replicative RNA strand 17
复制子 replicon 11
副黏病毒 paramyxovirus 277
副突变 paramutation 214
干扰素 interferon 329
肝素 heparin 115
冈崎片段 Okazaki fragment 10
隔裂基因 split gene 45
功能性互补 functional complementation 465
共翻译分泌 cotranslational secration 203
共整合体 cointegrate 88
共整合载体 cointegrate vector 494
光调节子 light regulatory element 220
光复活 photoreactivation 32
光复活酶 photoreactivation enzyme 32
光敏色素 phytochroine 219
光受体 photoreceptor 219
滚环模型 rolling circle model 12
合子 zygote 146
核骨架 nuclear scaffold 179
核壳体 nucleocapsid 277
核酶 ribozyme 185
核内小 RNA small nuclear RNA，snRNA 22
核内致癌蛋白 nuclear onco－protein 324
核片层 nuclear lamina 179
核受体 nuclear receptor 174
核糖核酸 ribonucleic acid，RNA 1
核糖体 ribosome 20
核糖体 RNA ribosomal RNA，rRNA 21
核外遗传 extranuclear inheritance 228
核小体 nucleosome 8，147
核小体重复 nucleosome repeat 147
核原生质配基 nucleoplasmin 206
恒定区 constant region 346
恒定域 constant domain 49
红霉素 erythromycin 125
后极系统 posterior systerm 420
后随链 lagging strand 10
呼肠孤病毒 reovirus 281
互变异构化 tautomerization 57
互补 DNA complementry DNA 17
互补决定区 complementarity－determining region，CDR 352
花斑发育 mosaic development 428
花椰菜花叶病毒 caulimovirus，CaMV 273
黄病毒 flavivirus 276
回文对称结构 palindrome 454
获得性免疫缺陷综合征 acquired immune deficiency syndrome 510
霍利迪连接体 Holliday junction 69
霍利迪模型 Holliday model 69
霍利迪中间体 Hollidoy intermediate 69
基因簇 gene cluster 53
基因家族 gene family 53
基因克隆 gene cloning 460
基因扩增 gene amplification 55
基因型 genotype 42
基因治疗 gene therapy 508
基因重排 gene rearrangement 351
基因转变 gene conversion 73
基因子 genelet 50
基因组 genome 10
基因组刻印 genome imprinting 213
激活子 activator 82
激素反应子 hormone response eleent 174
激肽释放酶 kallikrein 46

极性突变 polar mutation 93
脊髓灰质炎病毒 poliovirus 276
荚膜多糖 capsular polysaccharide 1
甲肝病毒 hepatitis 276
甲基化作用 methylation 93
甲基甲烷磺酸盐 methyl methane sulfonate 31
假基因 pseudogene 46
间叶细胞 mesenchyme cell 199
兼性异染色质 facultative heterochromatin 149
剪接体 spliceosome 185
剪接自动调节 splicing homeostasis 239
简并密码 degenerate codon 22
简并现象 degeneracy 23
碱基对 base pair, bp 6
碱基互补 base complementation 6
碱基类似物 base analog 61
交叉连接 cross link 31
交连桥前移 bridge migration 69
交配外激素 mating pheromone 47
交配型 mating type 404
交配型等位基因 mat allele 84, 405
胶原蛋白 collagen 53
焦磷酸转移酶 pyrophosphoryl transferase 140
铰链区 hinge 49
酵母人工染色体 yeast artificial chromosome 481
接合（配合）conjugation 112
拮抗信使 antimessenger 79
结构基因 structural gene 51
结合沟 binding groove 119
解离 dessociation 82
近侧序列元件 proximal sequense element 167
禁忌克隆 forbidden clone 341
精氨琥珀酸裂解酶 argininossuccinase 74
警报激素 alarmone 140
聚合酶链式反应 polymerase chain reaction, PCR 472
聚核糖体 polysome 28
均一着色区 homogeneously staining region 162
菌落杂交法 colony hybridiztion 463
卡拉霉素 kanamycin 77
卡隆载体 charon vector 479
抗体 antibody 49
抗体 antibody 49
抗原表位 epitope 344
抗原更换 antigenic shift 515
抗原决定簇 antigenic determinant 344
可变剪接 differential splicing 86
可变区 variable region 345
可变域 variable domain 49
可读框 open reading frame 42
克列诺片段 klenow fragment 105
刻印基因 imprinting gene 214
扩增控制因子 amplification control element 161
蓝光受体 blue－light receptor 219
劳斯氏肉瘤病毒 Rouas sarcoma virus 29
酪氨酰基 tyrosyl residue 8
类病毒 viroid 284
类固醇激素 steroid hormon 216
类固醇激素受体 steroid hormon receptor 216
利迪链菌素 streptolydigin 114
利福平 rifampicin 114
两性基因 double sex gene, dsx 213
亮氨酸拉链域 leucine zipper 175
淋巴母细胞样干扰素 lymphoblastoid interferon 330
流产启始期 abortive inituation 117

硫酸甲基甲烷 methylmethane sulfonate, MMS 63
侣伴蛋白 chaperon 206
侣伴蛋白宁 chaperonin 206
卵泡细胞 follecle cell 55
轮状病毒 rotavirus 281
螺旋酶 helicase 39, 103
螺旋体 solenoid 147
帽子结构 cap structure 163
帽子位点 cap site 43
密码子 codon 22
嘧啶二聚体 pyrimidine dimer 30
嘧啶碱 pyrimidine 4
免疫反应性 immunoreactivity 343
免疫记忆 immunological memory 340
免疫记忆细胞 immune memory cell 340
免疫监视 immune curveillance 340
免疫球蛋白 immunoglobulin, Ig 49
免疫球蛋白多样性 immunoglubulin diversity 357
免疫系统 immune system 341
免疫相关抗原 immune－associated antigen 368
免疫学 immunology 339
免疫原性 immunogenicity 343
免疫原性颗粒 immunogenic particle 514
模板链 template strand 17
末端蛋白 terminal protein 286
末端系统 terminal systerm 421
内部启动子 internal promotor 166
内含子 intron 45
内切核酸酶 endonuclease 452
内小核糖核蛋白颗粒 small nuclear ribonucleoprotein particle, snRNP 22
内域基因 internal gene 414
内源原病毒 endogenous provirus 297
拟病毒 virosoid 284
拟核 nucleoid 100
逆遗传学 reverse genetics 68
逆转录病毒 retrovirus 17
逆转录酶 reverse transcriptase 17
逆转录子 counter transcript 113
逆转录作用 reverse transcription 29
逆转座子 retrotransposon 85
黏性末端 sticky end 454
鸟嘌呤 guanine, G 4
鸟枪射击法 shotgun method 464
牛痘病毒 vaccinia virus 271
泡沫病毒 spumavirus 281
疱疹病毒 herpesvirus 270
配对保护模型 match protection model 243
普里布诺框 pribnow box 43
齐头末端 blunt end 454
启动点 start 434
启动子 promotor 43
启始因子 initiation factor, IF 26, 123
启始因子 initiation factor, IF 26, 123
前导链 leading strand 10
前导序列 leader 43
前基因 pregene 50
前极系统 anterior system 419
前酶原 preproenzyme 46
前启始复合体 preinitiation complex 170
前原引物复合体 prepropriming complex 109
前占竞争 preemptive competition 180
前—后轴 anterior－posterior 417
嵌入突变剂 intercalating agent 65
羟氨 hydroxylamine 31, 63
羟基化因子 hydroxylating agent 63
切割分裂 cleavage division 380
切割修复 excision repair 32
切口平移 nick translation 105
轻链 light strand, L 231

倾向错误修复 eror－prone repair 38
缺失 deletion 57，88
染色体步查法 chromosome walking 467
染色质 chromatin 8，146
热激蛋白 heat shock protein 141
人类免疫缺陷病毒 human immunodeficiency virus，HIV 328
绒毛膜促性腺激素 chorionic gonadotropin 54
溶解 melting 13
溶解酶 resolvase 78
融合蛋白 fusion protein 480
融合基因 chimeric gene 126
肉瘤 sarcoma 283
乳酸脱氢酶 lactate dehydrogenase，LDH 202
乳糖操纵元 lac operon 131
弱化反应 attenuation 130
弱化子，弱化基因 attenuator 134
扫描模型 scanning model 187
色氨酸操纵元 trp operon 133
杀伤性 T 细胞 killer T cell，Tk 343
上游活化序列 upstream activating sequense UAS 174
上游控制因子 upstream control element UCE 168
上游启动子 upstream promotor 166
上游因子 upstream element 43
神经生长因子 nerve growth factor 440
生成细胞 founder cell 431
噬菌粒 phagemid 477
首尾连接 tandem array 14
首先复制模型 replication－first model 88
首先重组模型 recombination－first model 88
受放线菌素 actinomycin 115
双病毒 geminivirus 275
双链断裂修复模型 double strand break repair model 74
双螺旋 double helix 6
双脱氧序列分析法 dideoxy sequencing method 469
顺反子 cistron 9
顺式作用元件 cis－acting elemnent 43
丝胶 sericin 200
丝裂霉素 C mytomycin C 31
丝纤蛋白 fibron 200
松弛型 DNA relaxed DNA form 8
松弛型菌株 relaxed strain 140
宿主整合因子 integration host factor，Ih 267
肽基位点 peptidyl site 21
肽基转移酶 peptidyl transferase 190
糖皮质激素 glucocorticoid 178
糖原合成酶 glycogen synthetase 202
天花病毒 smallpox virus 271
填充片段 stuffer fragment 479
跳跃基因 jumping gene 76
通读 readthrough 60
同感序列，共有序列 consesus sequense 43
同功 tRNA iso－accepting tRNA 23
同功酶 isoenzyme 202
同型排斥 isotype exclusion 356
同义突变 same sense mutation 23
同源异形 homoesis 423
同源异型基因 homeotic gene 193
同源异型框 homeobox 385
同源异型域 homeodomain 175
同宗配合 homothalism 404
拖尾 train 43
脱氨基因子 deaminating agent 63
脱辅基脂蛋白 apolipo protein－B，Apo－B 208
脱稳定子 destablizing element 195

脱氧核糖核酸 deoxyribonucleic acid，DNA 1
拓扑学反应 topological reaction 118
拓扑异构酶Ⅱ topoisomeraseⅡ 8
外激素 pheromone 411
外切核酸酶 exonuclease 452
外显子 exon 45
外遗传变异 epigenetic change 93
烷化剂 alkylating agent 63
挽回载体 retrieval vector 492
微管蛋白 tubulin 53
卫星 DNA satelite DNA 14，149
卫星病毒 satellite virus 280
位点专一性诱变 site－specific mutagenesis 65
无义链 nonsense strand 17
无义密码 nonsense codon 23
无义突变 nonsense mutation 58
误配对识别假说 mismatch recognitoin 243
误配对修正酶 mismatch correction enzyme 36
细胞癌基因 cellular onco－gene 317
细胞程序死亡 programmed cell death，或称 appoptosis 445
细胞毒性 T 细胞 cytotoxic T lymphocyte，CTL 或 Tc 343
细胞质定子 cytoplasmic determination 429
细胞质多角体病毒 cytoplasmic polyhedrosis virus 281
细胞质雄性不育 cytoplasmic male sterelity，CMS 236
细胞质遗传 cytoplasmic inheritance 228
细菌人工染色体 bacterial artificial chromosome，BAC 482
细小病毒 parvovirus 291
衔接重复 tandem repeat 54
显性控制区 dominant control region 166
显性位置效应花斑 dominant position effect variegation 214
限制性内切核酸酶 restriction enzymes 453
限制性片段长度多态性 restriction fragment lenth polymorphism，RFLP 455
限制性片断 restriction fragment 455
限制性作图 restriction mapping 456
线粒体 mitochondrion 56
腺病毒 adenovirus 9
腺嘌呤 adenine，A 4
相变 phase variation 403
小牛生长激素 bovin growth hormon BGH 126
心病毒 cardiovirus 276
锌指 zinc finger 175
锌指蛋白 zinc finger protein 172
信号识别颗粒 signal recognitin particle，SRP 203
信号肽 signal peptide 46
信号肽酶 signal peptidase 203
信使 RNA messenger RNA，mRNA 18
形态发生梯度 morphogenetic gradient 417
性致死基因 sex lethal gene 212
胸腺嘧啶 thymine，T 4
溴化乙锭 ethidiumbromide 65
选择内含子 optional intron 233
血红蛋白 hemoglobin 58
芽孢形成 sporation 400
亚基因组 mRNA subgenomic－messenger RNA 279
严紧反应 stringent response 139
叶绿体 chloroplast 56
胰岛素样生长因子 insulin－like growth vactor 214

遗传极性 genetic polarity 120
乙基甲烷磺酸盐 ethyl methane sulfonate 31
乙型肝炎病毒 hepatitis B virus 280
异染色质 heterochromatin 148
异宗配合 heterothalism 404
抑癌基因 tumor suppressor gene 322
抑制型操纵元 repressible operon 51
抑制性 T 细胞 suppressor T lymphocytes, STL 或 Ts 343
抑制子－突变子 suppressor－mutator 82
引物 primer 11
引物酶 primase 104
引物体 primosome 110
隐密基因 cryptogene 48
隐形蛋白基因 cryptomorphic gene 47
隐性突变 recessive mutation 76
有色体 chromoplast 146
有义链 sense strand 17
诱导型操纵元 inducible operon 51
鱼精蛋白 protamine 147
原癌基因 proto－oncogene 194
原病毒 provirus 282
原核 pronucleus 200
原黄素 proflavin 65
原脱植基叶绿素 protochlorophyllide 219
运输识别颗粒 transit recognition particle 46
运输肽 transit peptide 46
载体 vector 451
再启始位点 reinitiation site 126
早早期基因 immediate early gene 175
增变基因 mutator gene 60
增变菌株 mutator 60
增强子 enhancer 43
增殖细胞核抗原 proliferating cell nuclear antigen, PCNA 155
蔗糖酶 invertase 81
正超螺旋 positive supercoiling 8
植物生长素 auxin 217
中度重复 DNA middle repetitive DNA 14
中间载体 intermediate vector 494
中期骨架 metaphase scaffold 152
中枢免疫器官 central immune organ 341
中性突变 neutral mutation 59
终止密码 termination codon 23
终止位点 terminus 11
终止子 terminator 45
重叠基因 overlapping gene 56
重链 heavy strand, H 231
重组 recombination 68
重组信号片段 recombination signal sequence, RSS 356
重组修复 recombinational repair 33
珠蛋白 globin 28
主要组织相容性复合体 major histocompatibility complex, MHC 367
转变极性 conversion polarity 74
转变启始位点 conversion initiation site 74
转变因素 transforming principle 3
转化 transformation 3, 460
转化基因 transformer, tra 211
转换 transition 57
转甲基酶 transmethylase 138
转录 transcription 17
转录单位 transcriptional unit 42
转录活性染色体 transcriptionally active chromosome, TAC 251
转录降低调节子 transcriptional down regulator 179
转录泡 transcription bubble 117
转录因子 transcription factor, TF 170
转位 transvection 214
转移 RNA transfer RNA, tRNA 18
转移抑制基因 metastasis suppressor gene 323

转座酶 transposase 82
转座因子 transposable element 76
转座子 transposon 60
着色性干皮病 xeroderma pigmentosum XP 31
着丝点序列 centromeric sequense 149
着丝粒 centromere 149
自主因子 autonomous element 81
组成型操纵元 constitutive operon 51
组成型异染色质 constitutive heterochromatin 148
组蛋白 histone 146
组织相容性抗原 histocompatibility antigen 367
座位活化区 locus activating region 166
座位控制区 locus control region, LCR 166
吖啶 acridine 65

英文索引

abortive inituation 流产启始期 117
acquired immune deficiency syndrome 获得性免疫缺陷综合征 510
acridine 吖啶 65
actinomycin 受放线菌素 115
activator 激活子 82
adenine，A 腺嘌呤 4
adenovirus 腺病毒 9
alarmone 警报激素 140
alkylating agent 烷化剂 63
allo－lactose 别乳糖 132
Alu sequense Alu 序列 151
amino acyl－tRNA synthetase 氨酰基－tRNA 合成酶 18，134
amplification control element 扩增控制因子 161
amyloplast 淀粉体 247
aninoacyl site 氨酰基位点 21
anterior system 前极系统 419
anterior－posterior 前—后轴 417
anthranilate synthetase 133
antibody 抗体 49
anticodon 反密码子 18，22
antigenic determinant 抗原决定簇 344
antigenic shift 抗原更换 515
antigenome 反基因组 280
antimessenger 拮抗信使 79
antisenseRNA 反义 RNA 113
antiterminator 反终止子 120
aphtho virus 口疮病毒 276
apolipo protein－B，Apo－B 脱辅基脂蛋白 208
apurinic and apyrimidinic site AP 位点 31
argininossuccinase 精氨琥珀酸裂解酶 74
attenuation 弱化反应 130
attenuator 弱化子，弱化基因 134
autonomous element 自主因子 81
auxin 植物生长素 217
B－cell receptor editing B 细胞受体编排 359
bacterial artificial chromosome，BAC 细菌人工染色体 482
base analog 碱基类似物 61
base complementation 碱基互补 6
base pair，bp 碱基对 6
binary vector 二元载体 494
binding groove 结合沟 119
blue－light receptor 蓝光受体 219
blunt end 齐头末端 454
bovin growth hormon BGH 小牛生长激素 126
branch point sequense 分支点序列 184

bridge migration 交连桥前移 69
bypass synthesis 附属合成 38
cap site 帽子位点 43
cap structure 帽子结构 163
capsular polysaccharide 荚膜多糖 1
carcinoma 癌 283
cardiovirus 心病毒 276
catabolite activator protein, CAP 分解代谢活化蛋白 130
caulimovirus, CaMV 花椰菜花叶病毒 273
CCAAT box transcription factor, CTF CCAAT 框转录因子 173
cellular onco－gene 细胞癌基因 317
central immune organ 中枢免疫器官 341
centromere 着丝粒 149
centromeric sequense 着丝点序列 149
chaperonin 侣伴蛋白宁 206
chaperon 侣伴蛋白 206
charon vector 卡隆载体 479
chimeric gene 融合基因 126
chloroplast 叶绿体 56
chorionic gonadotropin 绒毛膜促性腺激素 54
chromatin 染色质 8，146
chromoplast 有色体 146
chromosome walking 染色体步查法 467
cis －acting elemnent 顺式作用元件 43
cistron 顺反子 9
cleavage division 切割分裂 380
closed circular DNA 闭合环状 DNA 100
coding sequense 编码序列 18
codon 密码子 22
cointegrate vector 共整合载体 494
cointegrate 共整合体 88
collagen 胶原蛋白 53
colony hybridiztion 菌落杂交法 463
complementary determining region, CDR 互补决定区 352
complementry DNA 互补 DNA 17
complex gene 复合基因 48
conjugation 接合（配合） 112
consesus sequense 同感序列，共有序列 43
constant domain 恒定域 49
constant region 恒定区 346
constitutive heterochromatin 组成型异染色质 148
constitutive operon 组成型操纵元 51
conversion initiation site 转变启始位点 74
conversion polarity 转变极性 74
coordinate control 并列控制 51
cotranslational secration 共翻译分泌 203
counter transcript 逆转录子 113
cross link 交叉连接 31
cryptogene 隐密基因 48
cryptomorphic gene 隐形蛋白基因 47
cytocine，C 胞嘧啶 4
cytoplasmic determination 细胞质定子 429
cytoplasmic inheritance 细胞质遗传 228
cytoplasmic male sterelity, CMS 细胞质雄性不育 236
cytoplasmic polyhedrosis virus 细胞质多角体病毒 281
cytotoxic T lymphocyte, CTL 或 Tc 细胞毒性 T 细胞 343
deaminating agent 脱氨基因子 63
degeneracy 简并现象 23
degenerate codon 简并密码 22
deletion 缺失 57，88
deoxyribonucleic acid, DNA, 脱氧核糖核酸 1
dessociation 解离 82
destablizing element 脱稳定子 195
developmental timer, DT 发育定时钟 378

dideoxy sequencing method 双脱氧序列分析法 469
differential splicing 可变剪接 86
differential gene expression 差别基因表达 214
dihydrofolate reductase 二氢叶酸还原酶 161
diplomorphic gene 二型蛋白基因 46
dispersed array 分散排列 14
dissociation 变性 13
DNA glycosylase DNA 糖基化酶 34
DNA gyrase DNA 回旋酶 8
DNA ligase DNA 连接酶 10
DNA polymerase DNA 聚合酶 10
DNA polymerase I DNA 聚合酶 I 11
DNA polymerase Ⅲ DNA 聚合酶Ⅲ 104
DNA recombination DNA 重组 451
DNA sequensing DNA 序列分析 468
DNA－binding domain DNA 结合域 175
dominant control region 显性控制区 166
dominant position effect variegation 显性位置效应花斑 214
dorsal－ventral systerm 背腹极性系统 421
dorsoventral 背腹轴 417
double helix 双螺旋 6
double sex gene, dsx 两性基因 213
double strand break repair model 双链断裂修复模型 74
dynamic competition 动力竞争 180
endogenous provirus 内源原病毒 297
endonucleases 内切核酸酶 452
enhancer 增强子 43
enterovirus 肠病毒 276
epigenetic change 外遗传变异 93
episome 附加体 112
epitope 抗原表位 344
Epstein Barr Virus EB 病毒 325
eror－prone repair 倾向错误修复 38
erythromycin 红霉素 125
ethidiumbromide 溴化乙锭 65
ethyl methane sulfonate 乙基甲烷磺酸盐 31
euchromatin 常染色质 148
excision repair 切割修复 32
exonucleases 外切核酸酶 452
exon 外显子 45
expression library 表达文库 465
expression vector 表达载体 483
extranuclear inheritance 核外遗传 228
facultative heterochromatin 兼性异染色质 149
fibroblast interferon 成纤维细胞干扰素 330
fibron 丝纤蛋白 200
flanking sequense 侧翼序列 10
flavivirus 黄病毒 276
follecle cell 卵泡细胞 55
forbidden clone 禁忌克隆 341
founder cell 生成细胞 431
functional complementation 功能性互补 465
fusion protein 融合蛋白 480
geminivirus 双病毒 275
gene amplification 基因扩增 55
gene cloning 基因克隆 460
gene cluster 基因簇 53
gene conversion 基因转变 73
gene family 基因家族 53
gene rearrangement 基因重排 351
gene therapy 基因治疗 508
geneamplification 基因扩增 158
genelet 基因子 50
genetic polarity 遗传极性 120

genome imprinting 基因组刻印 213
genome 基因组 10
genotype 基因型 42
globin 珠蛋白 28
glucocorticoid 糖皮质激素 178
glycogen synthetase 糖原合成酶 202
gratuitous inducer 安慰诱导物，义务诱导物 133
group 1 intron Ⅰ组内含子 240
group 2 intron Ⅱ组内含子 240
guanine，G 鸟嘌呤 4
half chromatid chiasma 半染色单体交叉 69
half life 半衰期 194
heat shock protein 热激蛋白 141
heavy strand，H 重链 231
helicase 螺旋酶 39，103
helper T lymphocyte，TH 辅助性 T 细胞 296
hemoglobin 血红蛋白 58
heparin 肝素 115
hepatitis B virus 乙型肝炎病毒 280
hepatitis 甲肝病毒 276
herpesvirus 疱疹病毒 270
heterochromatin 异染色质 148
heterogeneous nuclear RNA，hnRNA 不均一核 RNA 18
heterothalism 异宗配合 404
hinge 铰链区 49
histocompatibility antigen 组织相容性抗原 367
histone 组蛋白 146
Holliday junction 霍利迪连接体 69
Holliday model 霍利迪模型 69
Hollidoy intermediate 霍利迪中间体 H 69
homeobox 同源异型框 385
homeodomain HD 同源异型域 175
homeotic gene 同源异型基因 193
homoesis 同源异形 423
homogeneously staining region 均一着色区 162
homothalism 同宗配合 404
hormone response eleent 激素反应子 174
human immunodeficiency virus，HIV 人类免疫缺陷病毒 328
Hu factor Hu 因子
hydroxylamine 羟氨 31，63
hydroxylating agent 羟基化因子 63
hypoxanthin 次黄嘌呤 63
imcompatibility 不相容性 113
immediate early gene 早早期基因 175
immune curveillance 免疫监视 340
immune memory cell 免疫记忆细胞 340
immune system 免疫系统 341
immune－associated antigen 免疫相关抗原 368
immunogenic particles 免疫原性颗粒 514
immunogenicity 免疫原性 343
immunoglobulin，Ig 免疫球蛋白 49
immunoglubulin diversity 免疫球蛋白多样性 357
immunological memory 免疫记忆 340
immunology 免疫学 339
immunoreactivity 免疫反应性 343
imprinted gene 被刻印基因 214
imprinting gene 刻印基因 214
inducible operon 诱导型操纵元 51
initiation factor，IF 启始因子 26，123
insertiomal sequense 插入序列 76
insertional element 插入因子 60
insertion 插入 57
insulin－like growth vactor 胰岛素样生长因子 214
integration host factor，Ihf 宿主整合因子 267

intercalating agent 嵌入突变剂 65
interferon 干扰素 329
intermediate vector 中间载体 494
internal gene 内域基因 414
internal promotor 内部启动子 166
intron 内含子 45
inversion 倒位 88
invertase 蔗糖酶 81
inverted repeat 颠倒重复 54
IS insertional sequense IS 因子 77
iso－accepting tRNA 同功 tRNA 23
isoenzyme 同功酶 202
isotype exclusion 同型排斥 356
jumping gene 跳跃基因 76
kallikrein 激肽释放酶 46
kanamycin 卡拉霉素 77
killer T cell，Tk 杀伤性 T 细胞 343
kinetoplastid protozoa 动基体原生动物 236
klenow fragment 克列诺片段 105
lac operon 乳糖操纵元 131
lactamaseβ－酰胺酶 78
lactate dehydrogenase，LDH 乳酸脱氢酶 202
lagging strand 后随链 10
large T antigen 大 T 抗原 287
leader 前导序列 43
leading strand 前导链 10
Leaking mutation 泄漏突变 93
leucine zipper 亮氨酸拉链域 175
leukocyte interferon 白细胞干扰素 330
light regulatory element 光调节子 220
light strand，L 轻链 231
locus activating region 座位活化区 166
locus control region，LCR 座位控制区 166
long terminal repeat，LTR 长末端重复 283
lymphoblastoid interferon 淋巴母细胞样干扰素 330
M phase kinase M 期激酶 434
major grove 大沟 7
major histocompatibility complex，MHC 主要组织相容性复合体 367
mat allele 交配型等位基因 mat 84，405
match protection model 配对保护模型 243
maternal contributing cell type M 细胞型 86
mating pheromone 交配外激素 47
mating type 交配型 404
maturation promoting factor 成熟促进因子 434
melting 熔解 13
mesenchyme cell 间叶细胞 199
messenger RNA，mRNA 信使 RNA 18
metaphase scaffold 中期骨架 152
metastasis suppressor gene 转移抑制基因 323
methyl methane sulfonate 甲基甲烷磺酸盐 31
methylation 甲基化作用 93
methylmethane sulfonate，MMS 硫酸甲基甲烷 63
middle repetitive DNA 中度重复 DNA 14
mismatch correction enzyme 误配对修正酶 36
mismatch recognitoin 误配对识别假说 243
missense mutation 错义突变 58
mitochondrion 线粒体 56
monoclonal antibody 单克隆抗体 348
morphogenetic gradient 形态发生梯度 417
mosaic development 花斑发育 428
Mu phage Mu 噬菌体 80
multicopy plasmid 多拷贝质粒 112
muscle α－actin geneα－肌动蛋白基因 46
mutator gene 增变基因 60

mutator 增变菌株 60
mytomycin C 丝裂霉素 C 31
N protein N 蛋白 396
N－glycosylase 糖基化酶 31
negative super coiling 负超螺旋 8
nerve growth factor 神经生长因子 440
neutral mutation 中性突变 59
nick translation 切口平移 105
non－Mendelian inheritance 非孟德尔遗传 228
nonautonomous element 非自主因子 81
nonsense codon 无义密码 23
nonsense mutation 无义突变 58
nonsense strand 无义链 19
nuclear lamina 核片层 179
nuclear onco－protein 核内致癌蛋白 324
nuclear receptor 核受体 174
nuclear scaffold 核骨架 179
nucleocapsid 核壳体 277
nucleoid 拟核 100
nucleoplasmin 核原生质配基 206
nucleosome repeat 核小体重复 147
nucleosome 核小体 8，147
nucloid body 拟核体 236
ochre mutation 琥珀突变 78
Okazaki fragment 冈崎片段 10
oncovirus 癌病毒 281
open reading frame 可读框 42
operator 操作子 52
operon 操纵元 51
optional intron 选择内含子 233
overlapping gene 重叠基因 56
palindrome 回文对称结构 454
paramutation 副突变 214
paramyxovirus 副黏病毒 277
parvovirus 细小病毒 291
paternal contributing cell type P 细胞型 86
peptidyl site 肽基位点 21
peptidyl transferase 肽基转移酶 190
phagemid 噬菌粒 477
phase variation 相变 403
pheromone 外激素 411
photoreactivation enzyme 光复活酶 32
photoreactivation 光复活 32
photoreceptor 光受体 219
phytochroine 光敏色素 219
point mutation 点突变 57
polar efrfect 极性效应 93
polar mutation 极性突变 93
poliovirus 脊髓灰质炎病毒 276
poly A ribonuclease，PAN 多聚 A 核酸酶 195
poly A tail 多聚 A 尾巴 18
polyA polymerase 多聚 A 聚合酶 183
polycistronic mRNA 多顺反子 mRNA 18
polymerase chain reaction，PCR 聚合酶链式反应 472
polyomavirus 多瘤病毒 270
polysome 聚核糖体 28
positive supercoiling 正超螺旋 8
posterior systerm 后极系统 420
postitional cloning 定位克隆 467
poxvirus 痘病毒 270
preemptive competition 前占竞争 180
pregene 前基因 50
preinitiation complex 前启始复合体 170
preproenzyme 前酶原 46
prepropriming complex 前原引物复合体 109
pribnow box 普里布诺框 43
primase 引物酶 104
primer 引物 11
primosome 引物体 110
proflavin 原黄素 65

programmed cell death，或称 appoptosis 细胞程序死亡 445
proliferating cell nuclear antigen，PCNA 增殖细胞核抗原 155
promotor 启动子 43
pronucleus 原核 200
protamine 鱼精蛋白 147
protein release factor，RF 蛋白质释放因子 27
proto－oncogene 原癌基因 194
protochlorophyllide 原脱植基叶绿素 219
provirus 原病毒 282
proximal sequense element 近侧序列元件 167
pseudogene 假基因 46
psoralen 补骨脂素 31
purine 嘌呤碱 4
pyrimidine dimer 嘧啶二聚体 30
pyrimidine 嘧啶碱 4
pyrophosphoryl transferase 焦磷酸转移酶 140
pyruvate kinase 丙酮酸激酶 202
Q protein Q 蛋白 397
readthrough 通读 60
reassociation 复性 13
recessive mutation 隐性突变 76
recombination signal sequence，RSS 重组信号片段 356
recombination－first model 首先重组模型 88
recombinational repair 重组修复 33
recombination 重组 68
recruitment model 补给模型 118
regulative development 调节发育 428
regulatory gene 调节基因 51
regulon 调节元 51
reinitiation site 再启始位点 126
relaxed DNA form 松弛型 DNA 8
relaxed strain 松弛型菌株 140
reovirus 呼肠孤病毒 281
replication competent 复制胜任型 283
replication fork 复制叉 10
replication origin 复制启始区 11，152
replication－first model 首先复制模型 88
replicative form 复制型 274
replicative RNA strand 复制型 RNA 链 17
replicon 复制子 11
replisome 复制体 110
repressible operon 抑制型操纵元 51
resolvase 溶解酶 78
restriction enzymes 限制性内切核酸酶 453
restriction fragment lenth polymorphism，RFLP 限制性片段长度多态性 455
restriction fragment 限制性片断 455
restriction mapping 限制性作图 456
retrieval vector 挽回载体 492
retrotransposon 逆转座子 85
retrovirus 逆转录病毒 17
reverse genetics 逆遗传学 68
reverse transcription 逆转录作用 29
reverse transcroptase 逆转录酶 17
rhabdovirus 棒病毒 277
rhinovirus 鼻病毒 276
Ribolose－1，5－bisphosphate carboxylase/oxygenase，Rubisco Rubisco 260
ribonucleic acid，RNA，核糖核酸 1
ribosomal RNA，rRNA 核糖体 RNA 21
ribosome 核糖体 20
ribozyme 核酶 185
rifampicin 利福平 114
RNA editting RNA 编辑 23
RNA transcriptase RNA 转录酶 17
rolliag circle model 滚环模型 12

rotavirus 轮状病毒 281
Rouas sarcoma virus 劳斯氏肉瘤病毒 29
same sense mutation 同义突变 23
sarcoma 肉瘤 283
satelite DNA 卫星 DNA14，149
satellite virus 卫星病毒 280
scanning model 扫描模型 187
semi－conservative replication 半保留复制 10
sense strand 有义链 17
sericin 丝胶 200
sex lethal gene 性致死基因 212
Shine－Dalgarno sequense SD 序列 44
shotgun method 鸟枪射击法 464
signal peptidase 信号肽酶 203
signal peptide 信号肽 46
signal recognitin particle，SRP 信号识别颗粒 203
silencer 沉默子 165
single copy plasmid 单拷贝质粒 112
single stranded binding protein 单链结合蛋白 8
single－stranded DNA binding protein，SSB 单链 DNA 结合蛋白 8
site－specific mutagenesis 位点专一性诱变 65
small nuclear ribonucleoprotein particle，snRNP 内小核糖核蛋白颗粒 22
small nuclear RNA，snRNA 核内小 RNA 22
smallpox virus 天花病毒 271
solenoid 螺旋体 147
spliceosome 剪接体 185
splicing homeostasis 剪接自动调节 239
split gene 隔裂基因 45
sporation 芽孢形成 400
spumavirus 泡沫病毒 281
start 启动点 434
steroid hormon receptor 类固醇激素受体 216
steroid hormon 类固醇激素 216
sticky end 黏性末端 454
streptolydigin 利迪链菌素 114
stringent response 严紧反应 139
structural gene 结构基因 51
stuffer fragment 填充片段 479
subgenomic－messenger RNA 亚基因组 mRNA 279
supercoiling 超螺旋 8
suppressor T lymphocyte，STL 或 Ts 抑制性 T 细胞 343
suppressor－mutator 抑制子－突变子 82
supressor of RNA polymerase SRB RNA 聚合酶Ⅱ抑制子 170
T even bacteriophage T 偶列噬菌体 265
T odd bacteriophage T 奇数噬菌体 266
tagging 标签法 468
tandem array 首尾连接 14
tandem repeat 衔接重复 54
TATA binding protein，TBP TATA 序列结合的蛋白质 170
TATA box TATA 框 163
tautomerization 互变异构化 57
telomerase 端粒酶 156
telomere 端粒 150
telometric gene 端粒基因 414
template strand 模板链 17
terminal protein 末端蛋白 286
terminal systerm 末端系统 421
termination codon 终止密码 23
terminator 终止子 45
terminus 终止位点 11
thymine，T 胸腺嘧啶 4
T_n transposon 77
toll gene toll 基因 422

topoisomerase Ⅱ 拓扑异构酶Ⅱ 8
topological reaction 拓扑学反应 118
train 拖尾 43
trans－acting element 反作用子 52
trans－acting suppressor 反作用抑制子 83
transcription bubble 转录泡 117
transcription factor，TF 转录因子 170
transcriptional down regulator 转录降低调节子 179
transcriptional unit 转录单位 42
transcriptionally active chromosome，TAC 转录活性染色体 251
transcription 转录 17
transfer RNA，tRNA 转移 RNA 18
transformation 转化 3，460
transformer，tra 转化基因 211
transforming principle 转变因素 3
transit peptide 运输肽 46
transit recognition particle 运输识别颗粒 46
transition 转换 57
translation coupling 翻译偶联 126
translational operator 翻译操作子 126
translation 翻译 25
transmethylase 转甲基酶 138
transposable element 转座因子 76
transposase 转座酶 82
transposon 转座子 60
transvection 转位 214
transversion 颠换 57
trp operon 色氨酸操纵元 133
tubulin 微管蛋白 53
tumor suppressor gene 抑癌基因 322
tyrosyl residue 酪氨酰基 8
ubiquitin 泛素 203
unequal crossing over 不等交换 68
upstream element 上游因子 43
upstream activating sequense UAS 上游活化序列 174
upstream control element UCE 上游控制因子 168
upstream promotor 上游启动子 166
V－oncogene 病毒癌基因 316
vaccinia virus 牛痘病毒 271
variable domain 可变域 49
variable region 可变区 345
variable surface glycoprotein 表面糖蛋白变体 412
vector 载体 451
viroid 类病毒 284
virosoid 拟病毒 284
virus gene promotor 病毒基因启动子 302
virus transcriptional termination 病毒基因转录终止 306
virus 病毒 265
vitellogenin 卵黄前体蛋白 216
wobble hypothesis 摆动假说 23
xeroderma pigmentosum XP 着色性干皮病 31
yeast artificial chromosome 酵母人工染色体 481
zinc finger protein 锌指蛋白 172
zinc finger 锌指 175
zygote 合子 146
α－amanitine α－鹅膏蕈碱 115
λ bacterio phage λ噬菌体 267
σ－aminolevulinate synthase DALA σ－氨基乙酰丙酸合成酶 198

主要参考书

〔1〕朱玉贤，李毅编著．现代分子生物学．北京：高等教育出版社，1995
〔2〕吴乃虎编著．基因工程原理．北京：高等教育出版社，1989
〔3〕谢友菊编著．遗传工程概论．北京：北京农业大学出版社，1992
〔4〕瞿礼嘉等．现代生物技术导论．北京：高等教育出版社，施普林格出版社，1998
〔5〕阎隆飞，张玉麟主编．分子生物学．北京：中国农业大学出版社，1997
〔6〕张茜蘅，童坦君等．医学分子生物学．北京：北京医科大学出版社，1999
〔7〕余传霖，叶天星，陆德源．现代医学免疫学．上海：上海医科大学出版社，1998
〔8〕翁心华，潘效彰等．现代感染病学．上海：上海医科大学出版社，1998
〔9〕毕爱华等．医学免疫学．北京：人民卫生出版社，1994
〔10〕徐耀先，周晓峰，刘立德．分子病毒学．武汉：湖北科学技术出版社，2000
〔11〕侯云德等．分子病毒学．北京：学院出版社，1990
〔12〕吴逸明，安玉会．基础肿瘤学．郑州：河南医科大学出版社，1998
〔13〕谷志远，赵亚力．现代医学分子生物学．北京：人民军医出版社，1998
〔14〕翁心华，潘效彰．现代感染病学．上海：上海医科大学，1998
〔15〕董长亘，吴时友．现代分子病毒学．武汉：武汉大学出版社，1996
〔16〕张茜蘅，童坦君等．医学分子生物学．北京：北京医科大学出版社，1999：775～869
〔17〕余传霖，叶天星，陆德源．现代医学免疫学．上海：上海医科大学出版社，1998：168～396
〔18〕翁心华，潘效彰等．现代感染病学．上海：上海医科大学出版社，1998：3～275
〔19〕毕爱华等．医学免疫学．北京：人民卫生出版社，1994
〔20〕翟中和主编．细胞生物学．北京：高等教育出版社，1998
〔21〕吴克中编著．细胞生长调节因子．北京：科学出版社，1988
〔22〕刘景生主编．细胞信息与调控．北京：北京医科大学，中国协和医科大学联合出版社，1998
〔23〕丁汉波，仝允栩，黄淅．发育生物学．北京：高等教育出版社，1987
〔24〕陈吉龙，马海飞主编．发育生物学进展．北京：高等教育出版社，1994
〔25〕陈汉源．p53 蛋白质结构和生物功能．国外医学遗传分册，18:(5)1239～242,1995
〔26〕孙乃恩，孙东旭，朱德煦编著．分子遗传学．南京：南京大学出版社，1993

〔27〕李汝琪．发育遗传学．上、下册．北京：科学出版社，1985

〔28〕王瑞虹，薛绍白．高等真核细胞 Pr. 磷酸化与细胞周期 Gn 向 S 期过渡的调控．细胞生物学，(2)：71～73，1995

〔29〕Sambrook J.，Fritsch E.F.，Maniatis T.，Molecular Cloning. Cold Spring Harbor Laboratory Press 1989

〔30〕Robert F.Weaver，Philip W.Hedrick.Genetics.Wm.C.Brown Publishers.USA.1997

〔31〕Anthony J.F.Griffths，Jeffrey H.Miller，David T.Suzuki，Richard C.Lewontin，William M.Gelbart. Genetic Analysis.WH Freeman and Company.England.1996

〔32〕Sassone－Corsi，P.Molecular Clocks：Inastering time by gone regulation Natrue，V392：30：871～872

〔33〕Strager，P and losick，R.Cascades of sigma factors revisited molecular Microbiology 4 (11)：1801～1806，1990

〔34〕Lodish，H.Baltimore，D.Berk，A.，S.L.Zipursky，P.Matsudaira，J.Darnell，Molecular Cell Biology，W.H.Freeman and Compaing，New York，1995

〔35〕Nusse，R.Varmus，H.Wnt Genes Cell，69：1073～1087，1993

〔36〕Watson，J.D.，N.H.Hopkins，J.W.Roberts，J.A.Steitzand A.M.Weiner，Molecular Biology of the Gene.4th eddition，New York，1987

〔37〕Dillon，L.S.（ed.），The Gene－its structure function are Evolution，Plenum Press，New Yourk.1986

〔38〕Ketchum，P.A.，Microbiology－Concepts and Apllication，John Wiley and sone，New York.1985

〔39〕Mahy，B.W.J.Virology－a Practical Approach，IRL Prss，Oxford，England.1985

〔40〕Schleif，R.Genetics and Molecular Biology，Addison－Wesley Pub.USA.1986

〔41〕Taylor，W.C.Regulatory Interactions Between Nuclear and Plastid genomes，Ann.Rev.Plant Physiol.Plant Mol.Biol.40：211～233.1989

〔42〕Schlegel，R.A.，M.S.Halleck and P.N.Ras（eds）Molecular Regulation of Nuclear Events in Mitosis and Meiosis，Aeadmic Press INC.，Orlando.1987

〔43〕Lewin，B.，Genes Ⅶ，Oxford Univ.Press.2000

〔44〕Weaver，R.F.and P.W.Hedrick，Genetics 3rd edition，WCB Publisher，London.1997

〔45〕Griffiths，A.J.F.，J.H.Miller，D.T.Suzuki，R.C.Lewontin and W.M.Gelbart An Introduction to Genetic Analysis W.H.Freeman and Company.New York.1996

[General Information]
书名=分子遗传学
作者=杨业华主编
页数=541
SS号=10346804
出版日期=2001年01月第1版
出版社=中国农业出版社
尺寸=23cm
原书定价=45.70
主题词=分子遗传学(学科: 高等学校)分子遗传学
参考文献格式=杨业华主编.分子遗传学.中国农业出版社,2001.
内容提要=面向21世纪课程教材:本书阐述了遗传信息的复制,基因的结构、功能和调节表达机理,基因突变与重组,基因对发育的控制以及基因操作及其应用的基本原理。

封面页
书名页
版权页
前言页
目录页
第一章　遗传物质ＤＮＡ和ＲＮＡ
第一节　ＤＮＡ分子的结构与复制
一、遗传物质ＤＮＡ的发现与证明
二、ＤＮＡ分子的结构
三、ＤＮＡ分子的功能单位
四、ＤＮＡ分子的复制
五、ＤＮＡ序列组成
第二节　ＲＮＡ的结构与功能
一、ＲＮＡ分子的结构
二、ＲＮＡ合成与病毒ＲＮＡ遗传信息传递
三、细胞中主要的ＲＮＡ
第三节　三联体密码与蛋白质合成
一、三联体密码
二、翻译
三、中心法则及其发展
第四节　ＤＮＡ损伤及其修复
一、ＤＮＡ变异
二、ＤＮＡ损伤的修复机制
小结
复习题
第二章　基因的特征
第一节　基因的结构特征
一、基因的一般结构
二、基因的变异类型
第二节　基因的组织
一、操纵元
二、重复基因与基因家族
三、重叠基因
第三节　基因突变与重组
一、基因突变
二、基因重组
三、基因转变
第四节　转座子
一、转座子的遗传特征
二、细菌中主要的转座子
三、植物转座子

四、酵母的转座子
五、无脊椎动物的转座子
六、转座子转座与基因突变
小结
复习题
第三章 原核生物基因组及其表达
第一节 细菌基因组及其复制
一、细胞中的拟核
二、细菌ＤＮＡ复制的主要酶类
三、细菌ＤＮＡ复制的模型
四、质粒的一般特征及其复制调节
第二节 细菌基因组的转录与翻译
一、与转录有关的酶
二、转录启始、延长与终止
三、翻译启始
第三节 操纵元的调节表达
一、操纵元调节表达的一般模型
二、乳糖操纵元的负控制和正控制诱导表达
三、色氨酸操纵元与弱化反应
四、核糖体蛋白操纵元与翻译自动调节
五、转录的总体控制
六、热激蛋白的合成
小结
复习题
第四章 真核生物细胞核基因组及其表达
第一节 ＤＮＡ与染色体
一、组蛋白与染色质
二、着丝粒与卫星ＤＮＡ
三、端粒
四、分散的高度重复ＤＮＡ
五、细胞分裂中期的骨架
第二节 细胞核ＤＮＡ复制
一、复制子及复制启始区
二、ＤＮＡ复制的蛋白质或酶类
三、细胞核ＤＮＡ复制模型
四、染色体端粒ＤＮＡ的合成
五、有选择的ＤＮＡ复制与基因扩增
第三节 基因的转录与蛋白质合成
一、转录酶
二、转录的控制信号
三、转录启始及转录因子

四、mRNA的加工
五、蛋白质合成的启始
六、翻译延长和终止
第四节 细胞核基因表达的控制
一、基因表达在转录和转录后水平上的控制
二、基因表达在翻译水平上的调节
三、基因表达的翻译后调节
第五节 发育阶段对基因表达的调节
一、基因表达的细胞、组织和器官特异性
二、脊柱动物珠蛋白基因表达的发育阶段调节
三、性别对基因表达的调节
第六节 环境信号对基因表达的调节
一、半乳糖对酵母基因表达的调节
二、外源激素对基因表达的调节
三、热激效应对基因表达的调节
四、光照对植物基因表达的调节
五、缺氧对植物基因表达的调节
小结
复习题
第五章 线粒体和叶绿体基因组的组织及表达
第一节 线粒体基因组
一、线粒体DNA的复杂性
二、哺乳动物线粒体基因组
三、酵母线粒体基因组
四、植物线粒体基因组
五、原生动物线粒体基因组
六、线粒体基因密码的不通用性
七、线粒体基因组的复制
第二节 线粒体基因的转录与翻译
一、线粒体基因的转录和转录控制
二、线粒体转录子的加工
三、RNA编辑
四、线粒体mRNA的翻译和翻译控制
第三节 植物质体基因组的组织及表达
一、质体基因组的组织和结构
二、质体基因的表达
第四节 线粒体和质体基因组与细胞核基因组的互作
一、细胞器的生物发生与核基因组的关系
二、细胞溶质蛋白向细胞器的运输
三、细胞器基因与核基因表达的协调性
四、植物线粒体与细胞质雄性不育性

小结
复习题
第六章　病毒基因组的表达及病毒感染对宿主细胞的遗传效应
第一节　病毒基因组的组织
一、双链ＤＮＡ噬菌体基因组
二、哺乳动物双链ＤＮＡ病毒基因组
三、植物双链ＤＮＡ病毒基因组
四、单链ＤＮＡ病毒基因组
五、ＲＮＡ病毒基因组
六、逆转录病毒基因组
七、植物类病毒和拟病毒基因组
第二节　病毒基因组的复制
一、线状双链ＤＮＡ病毒基因组的复制
二、环状双链ＤＮＡ病毒基因组的复制
三、单链ＤＮＡ病毒基因组的复制
四、单链ＲＮＡ病毒基因组的复制
五、逆转录病毒基因组的复制
第三节　病毒基因的转录及蛋白质合成
一、病毒基因的转录启始信号
二、病毒基因的转录终止信号
三、ＲＮＡ病毒蛋白质合成的策略
第四节　病毒与肿瘤
一、肿瘤及癌的一般生物学概念
二、动物及人体的致癌因子
三、癌基因、抑癌基因及肿瘤转移调节基因
四、ＤＮＡ肿瘤病毒与癌
五、ＲＮＡ病毒与肿瘤
第五节　病毒感染与宿主干扰素的合成
一、干扰素
二、干扰素基因
三、干扰素合成
四、干扰素的抗病毒活性
小结
复习题
第七章　脊椎动物免疫的基因控制
第一节　脊椎动物的免疫反应
一、免疫的概念
二、免疫系统与免疫反应
三、抗原及抗原性
第二节　Ｂ细胞免疫反应
一、抗体、单克隆抗体及其基因工程抗体

二、抗体基因及其重组
三、抗体基因重排与抗体多样性
第三节　T细胞免疫反应
一、T细胞的类型及对外源抗原的识别
二、T细胞受体基因
三、TCR基因的重排及其多样性产生的机制
第四节　主要组织相容性复合体
一、小鼠主要组织相容性复合体
二、人类MHC-HLA复合体
三、MHC的生物学意义
小结
复习题
第八章　基因对发育和细胞分化的控制
第一节　基因控制发育的方式
一、发育阶段的时间决定
二、基因控制发育的方式
三、Wnt基因家族与发育
第二节　λ噬菌体的发育调节
一、λ噬菌体生长概况
二、CI抑制子和Cro蛋白的作用机制
三、CII蛋白与噬菌体发育
四、整合酶基因表达的逆调节
五、N和Q蛋白的反终止作用对λ噬菌体发育的正控制
六、噬菌体DNA复制的特殊启动因子
七、单位λ基因组的形成
八、λ噬菌体裂解周期的诱导
第三节　细菌发育的基因控制
一、细菌芽孢形成的基因控制
二、鼠沙门氏杆菌表面抗原的转换
第四节　酵母交配型的转换
一、酵母的生活周期
二、酵母的快速性转变
三、酵母交配型盖合转换的调节
四、mat　a和mat　a基因编码控制蛋白
五、外激素与单倍体细胞的交配感受性
第五节　锥虫表面抗原转换的基因控制
一、锥虫生活史及表面抗原转换现象
二、锥虫表面抗原转换的机理
第六节　高等真核生物发育阶段的基因控制
一、果蝇中编码体型决定因子的基因
二、果蝇的toll基因与形态发生梯度

三、果蝇器官分化的同源异型基因
四、分化的反式决定
五、细胞质定子
六、植物花芽分化的同源异型基因
第七节　基因对细胞周期的控制
一、细胞周期事件与基因表达
二、细胞生长的控制
三、细胞程序死亡
小结
复习题
第九章　基因操作技术及其应用
第一节　ＤＮＡ重组的基本技术
一、核酸酶和限制性核酸内切酶与ＤＮＡ分子的体外切割
二、限制性片段长度多态性
三、限制片段作图
四、限制性片段的连接与克隆
五、ＤＮＡ序列分析
六、聚合酶链式反应
第二节　基因操作的载体系统
一、质粒载体
二、单链噬菌体克隆载体
三、噬菌粒－－含有丝状噬菌体复制起始位点的质粒
四、双链噬菌体载体
五、黏粒载体
六、酵母人工染色体和细菌人工染色体
七、表达载体
第三节　酵母的遗传工程
一、酵母载体
二、酵母载体的应用
第四节　植物的遗传转化
一、转基因植物
二、植物基因操作的研究方向与重要基因的鉴定
第五节　动物的遗传转化
一、动物转基因的方法
二、转化体的鉴定和分析
三、重组ＤＮＡ分子导入受体细胞
四、动物基因工程的发展前景
五、动物体细胞克隆技术
第六节　重组ＤＮＡ技术在医学上的应用
一、基因治疗
二、爱滋病的治疗

三、生物技术新药
四、基因工程与疫苗生产
五、重组ＤＮＡ探针与遗传病诊断
小结
复习题
中文索引
英文索引
主要参考书
附录页